

**DEVELOPMENTAL
NEUROTOXICITY STUDIES**

Standard Evaluation Procedure

Health Effects Division

March, 2009

TABLE OF CONTENTS

LIST OF TABLES.....	4
I. INTRODUCTION.....	5
A. <i>Background</i>	5
B. <i>Using this Standard Evaluation Procedure (SEP)</i>	5
II. DATA ACCEPTABILITY	6
A. <i>Acceptance Criteria</i>	6
B. <i>Study Design</i>	7
III. EVALUATION OF STUDY CONDUCT	7
A. <i>Test Compound</i>	7
B. <i>Animal Selection</i>	7
1. Species	7
2. Number of Animals	7
3. Assignment of animals for behavioral tests, brain weights, and neuropathological evaluations	9
C. <i>Dose Selection and Dosing</i>	9
1. Dose Selection	9
2. Dosing	10
a. Administration of the test article	10
b. Dosing Period	10
c. Verification of Exposure to Pups	10
D. <i>Maternal Observations</i>	11
1. General	11
2. Field Observation Battery (FOB)	11
E. <i>Litter Observations</i>	11
F. <i>Offspring Observations</i>	12
1. Physical Development	12
2. Post Weaning Observations	12
3. Neurobehavioral Evaluations	12
a. FOB	12
b. Motor Activity	12
c. Auditory Startle Test	13
d. Learning and Memory Tests	13
4. Postmortem Evaluations	15
a. Brain Weight and Fixation	15
b. Neuropathology	16
i. Qualitative Analysis	16
ii. Subjective Analysis	17
iii. Simple morphometric analysis	17
G. <i>Biomarkers of Exposure</i>	18
H. <i>Statistical Evaluations</i>	18

IV.	STUDY INTERPRETATION	19
A.	<i>Endpoints of Maternal Toxicity</i>	19
1.	Mortality and Clinical Signs	19
2.	Maternal Body Weight	20
3.	Food and water consumption	20
4.	Reproductive Performance	22
5.	Organ Weight and Pathology	22
B.	<i>Endpoints of Offspring Toxicity</i>	23
1.	Viability and Clinical Signs	23
2.	Physical Development	24
a.	Body Weight	24
b.	Developmental Landmarks	26
3.	Neurobehavioral Evaluations	27
a.	Functional Observation Battery (FOB)	28
b.	Motor Activity	29
c.	Auditory Startle	32
d.	Learning and Memory	34
4.	Postmortem	38
a.	Brain weights	39
b.	Neuropathology	40
c.	Morphometrics	41
C.	<i>Statistical Analyses</i>	42
D.	<i>Use of Control Data</i>	44
1.	Concurrent (vehicle) Control	44
2.	Historical Control	44
3.	Positive Control	45
V.	DEVELOPMENTAL NEUROTOXICITY HAZARD CHARACTERIZATION	46
VI.	DATA EVALUATION REPORT	48
VII.	REFERENCES	48

LIST OF TABLES

Table 1. Functional Observation Battery for Maternal Animals.....	11
Table 2. Selected mean (\pm SD) Maternal Body weight, Body Weight Gain and Food Consumption	21
Table 3. Reproductive Performance	22
Table 4. Factors influencing offspring viability.....	23
Table 5. Litter Size and Viability	24
Table 6. Mean (\pm SD) Pre-weaning Pup Body Weights (g).....	26
Table 7. Mean (\pm SD) Post-Weaning Pup Body Weights (g)	26
Table 8. Mean (\pm SD) Age of Sexual Maturation (days) and	27
Table 9. Functional Observational Battery Results (Incidence)	29
Table 10. Mean (\pm SD) Motor Activity Data (Total Number of Movements/50 minutes)	31
Table 11. Mean (\pm SD) Sub-session Motor Activity in Males	31
Table 12. Mean (\pm SD) Overall (Blocks 1-5) Acoustic Startle Peak Amplitude (mv), Latency to Peak (msec) and Average Response Amplitude (mv)	33
Table 13. Mean (\pm SD) Interval Acoustic Startle Peak Amplitude (mv), Latency to Peak (msec) and Average Response Amplitude (mv) In Males.....	34
Table 14. Passive Avoidance Performance (mean \pm SD).	36
Table 15. Water Maze Percentage of Successful Trials.....	37
Table 16. Mean Percentage of Successful Trials at 1.5 \times Straight Channel Swim Time	38
Table 17. Brain (\pm SD) Brain Weights	40
Table 18. Mean (\pm SD) Morphometric data.....	42

I. INTRODUCTION

A. Background

Developmental neurotoxicity (DNT) refers to any adverse effect on the normal development of nervous system structure and/or function resulting from pre-natal and early post-natal exposure to a toxic substance on the normal development of nervous system structure and/or function. The U.S. Environmental Protection Agency (USEPA) and the Organization for Economic Cooperation and Development (OECD) have developed DNT test guidelines designed to generate data on the potential functional and morphologic hazards to the nervous system in offspring of mothers exposed during pregnancy and lactation. Offspring undergo assessments for physical development, behavioral ontogeny, motor activity, motor and sensory function, and learning and memory, and the evaluation of brain weights and neuropathology during postnatal development and adult hood (USEPA, 1998a; USEPA, 1999; OECD, 2007).

According to Part 158 Toxicology Data Requirements for Pesticides (158.500), a DNT is “Conditionally Required” for registration of both food-use and non-food use pesticide chemicals. A weight-of-evidence approach is utilized by which the DNT is required for a pesticide chemical when evidence of neurotoxicity is seen in the standard toxicity, developmental, reproductive, or neurotoxicity studies and/or when there are structure-activity concerns (CFR 40, 2008).

B. Using this Standard Evaluation Procedure (SEP)

- **This Standard Evaluation Procedure (SEP) is not intended as a single set of rules to follow in the evaluation of developmental neurotoxicity studies. This SEP should be used in conjunction with the following documents:**
 - HED. Standard Evaluation Procedure: Developmental Toxicity Studies (USEPA, 1993a);
 - HED Standard Evaluation Procedure: Reproductive Toxicity Studies (USEPA, 1993b);
 - Guidelines for Developmental Toxicity Risk Assessment (USEPA, 1991);
 - Guidelines for Reproductive Risk Assessment (USEPA, 1996);
 - Guidelines for Neurotoxicity Risk Assessment (USEPA 1998b);
 - Guidance Document for Neurotoxicity Testing (OECD, 2004); and
 - Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment (OECD, 2008)

The scope of this document is intended primarily to assist in the identification of possible adverse effects and the dose level at which these effects occur or are absent. The guidance pertains to interpretation of the endpoints evaluated within the context of this particular study, however, it should be recognized that in the overall evaluation of the hazard associated with a chemical, a weight of the evidence approach should be taken. Typically, this would involve consideration of all the effects seen in this study as well as the results of the other toxicity studies with the same chemical in other species and/or the outcome of studies with analogous chemicals. Additionally, this guidance is confined to the interpretation of effects observed within the defined experimental animal test system. The relevance of the endpoints evaluated and their appropriate extrapolation in assessing hazard to human health are not addressed with in this document.

II. DATA ACCEPTABILITY

A. Acceptance Criteria

A list of acceptance criteria for the DNT is given below. This list is guidance for both registrants and reviewers as a summary of the data most important to the acceptability of a DNT study. Failure to meet all criteria does not necessarily invalidate a study nor does meeting all criteria automatically validate a study. Each study should be evaluated on a case-by-case basis.

1. Technical form of the active ingredient tested is reported.
2. At least 20 litters per dose group were tested.
3. Rat is the preferred species; however, NOT Fischer 344 rats.
4. Dosing of dams was Gestation Day (GD) 6 through Post Natal Day (PND) 21.
5. The highest dose tested should produce overt maternal toxicity, but should not induce *in utero* or neonatal death sufficient to preclude meaningful evaluation.
6. The lowest dose tested should not produce maternal or developmental neurotoxicity.
7. Analysis of test material stability, homogeneity, and concentration in dosing medium is reported.
8. Daily cage side observations are documented.
9. Detailed clinical observations are reported at least twice during gestation and twice during lactation.
10. Individual maternal body weight is recorded weekly during gestation and on lactation days 0, 1, and 21.
11. Litters standardized to 4/sex (or 5/sex) on PND 4.
12. Individual offspring body weight is documented on PND 0, 4, 11, 17, 21, and every 2 weeks until termination.
13. Sexual maturation of the offspring monitored (vaginal opening, preputial separation) and documented.
14. Age-appropriate FOB on 10 offspring/sex/dose on PND 4, 11, 21, 35, 45, 60.
15. Motor activity testing on 10 offspring/sex/dose on PND 13, 17, 21, 60.
16. Auditory startle response on 10 offspring/sex/dose at weaning and PND 60.
17. Learning and memory testing on 10 offspring/sex/dose at weaning and PND 60.
18. Brain weight from 10 offspring/sex/litter on PND 11/21 and 60.
19. Brain fixation by immersion on PND 11 and by perfusion on PND 21 and 60.
20. Microscopic neuropathology on samples from all major brain regions from 10 offspring/sex on PND 11/21 and 60 from at least the control and high-dose groups.
21. Simple Morphometric analysis of the brain on PND 11/21 and 60 minimally including the neocortex, hippocampus, and cerebellum.
22. Positive control data which demonstrate the ability of the testing facility to perform DNT studies is included.

B. Study Design

The general study design is shown in Figure 1. In a DNT study, the test substance is administered to several groups of pregnant adult animals during gestation and lactation, one dose level being used per group. Detailed clinical observations of dams are conducted at least twice during gestation and twice during lactation. Offspring are randomly selected from within litters for neurotoxicity evaluation. The evaluation includes observations to detect gross neurologic and behavioral abnormalities, determination of motor activity, auditory startle, assessment of learning and memory, brain weight and neuropathological evaluation (qualitative and quantitative)

III. EVALUATION OF STUDY CONDUCT

A. Test Compound

A DNT study should be performed with the technical form of the active ingredient intended for commercial use. The specifications of the test material should be documented and the concentration of the active ingredient(s) should be clearly indicated in the study report. This information should also be included in the study evaluation for comparison with material utilized in other studies. If a vehicle is used, it should not produce any systemic, developmental, or neurotoxic effects. If there is any question as to the toxicity of the vehicle, the registrant is required to provide data on the toxicity of the vehicle and/or justify the choice of the vehicle.

C. Animal Selection

1. Species

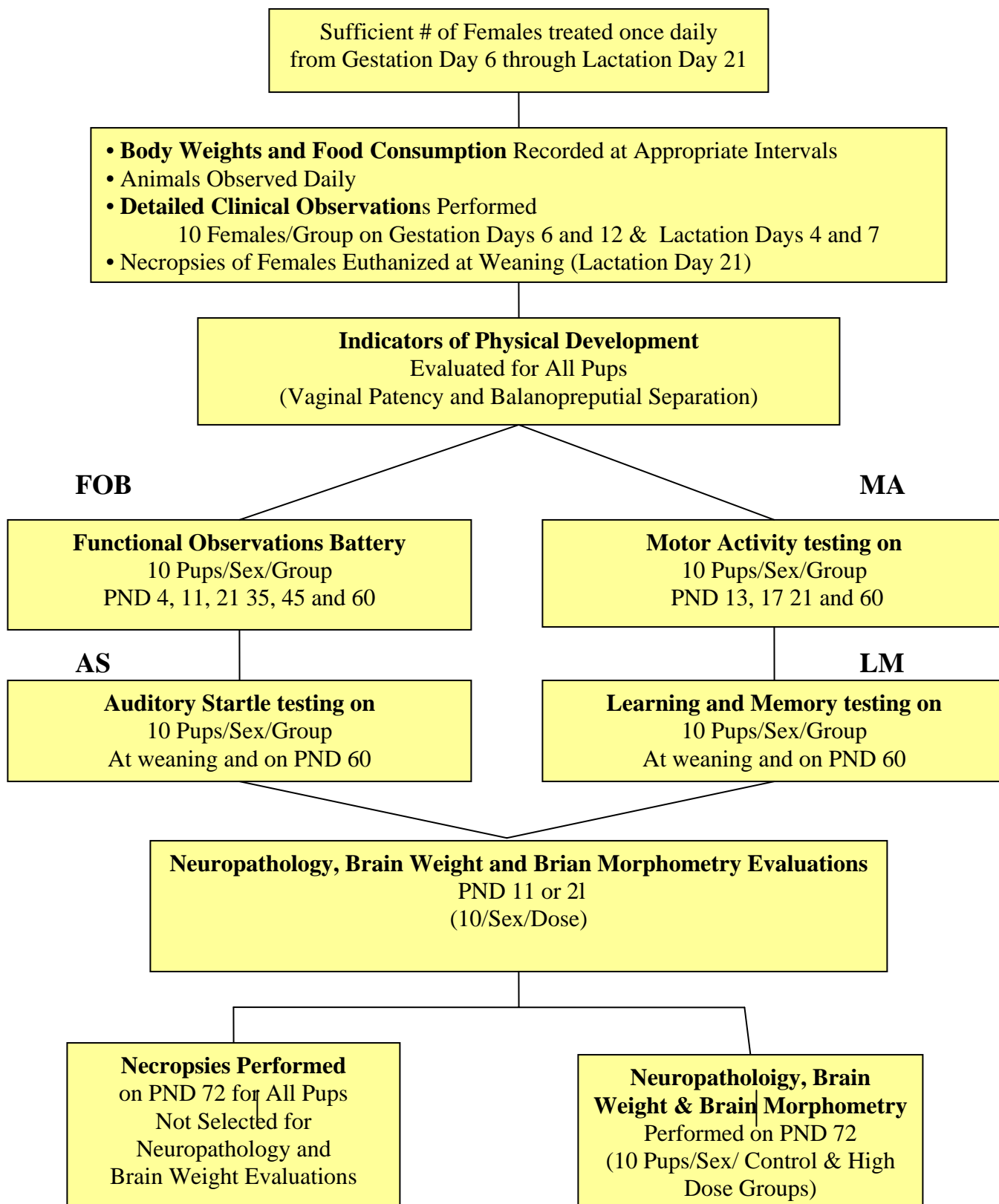
Testing should be performed in the rat. Young adult, nulliparous and pregnant females should be used at each dose level. Because of its differences in timing of developmental events compared to strains that are more commonly tested in other developmental and reproductive toxicity studies, it is preferred that the **Fischer 344 strain not be used**. If a sponsor chooses to use the Fischer 344 rat or a mammalian species other than the rat, ample justification/reasoning for this selection must be provided (USEPA 1998a).

2. Number of Animals

The objective of the study is for a sufficient number of pregnant rats to be exposed to the test substance to ensure that an adequate number of offspring are produced for neurotoxicity evaluation. Therefore, at least 20 litters per dose group are recommended.

On or before postnatal day (PND) 4 (day of delivery is PND 0), the size of each litter should be standardized by eliminating extra pups by random selection to yield, as nearly as possible, four males and four females per litter. Other litter size standardizations (e.g., 5/sex) are also acceptable. Whenever the number of pups of either sex prevents having four of each sex per litter, partial adjustment (e.g., five males and three females) is permitted. Standardization is not appropriate for litters of less than seven pups. Selective elimination of pups (e.g. runts based on body weights) is not appropriate. Individual pups should be uniquely identified after standardization of litters (USEPA 1998a).

Figure 1. Design of a Developmental Neurotoxicity Study



3. Assignment of Animals for Behavioral Tests, Brain Weights, and Neuropathological Evaluations

After standardization of litters, one male or one female from each litter (at least a total of 10 males and 10 females per dose group) should be randomly assigned to each of the following tests: motor activity, auditory startle, and learning and memory, in weanling and adult animals. Animals may be used for more than one test (e.g., the same set of animals may be used for FOB and motor activity evaluations). On PND 11 (or 21) either one male or one female pup from each litter (total of 10 males and 10 females/dose group) should be sacrificed or brain weights measured. An additional group of at least 10 animals per sex per group should be selected for neuropathological evaluation. If the assessment is conducted on PND 21, perfusion of the brain is required.

At the termination of the study, either one male or one female from each litter (total of 10 males and 10 females per dose group) should be sacrificed and brain weights measured. An additional group of at least 10 animals per sex per dose group (one male or one female per litter) should be sacrificed at the termination of the study for neuropathological evaluation performed in accordance with *Health Effects Test Guidelines OPPTS 870.6200 Neurotoxicity Screening Battery*.

The number of animals assigned for behavioral tests, brain weight measurements and neuro-pathological evaluations were increased from 6 sex/dose (as stated in the 870.6300 Guidelines) to 10/sex/dose in the Data Call In (DCI) issued by the Agency on September 10, 1999. In addition, Registrants were given the option of performing neuropathological evaluations at PND 21 instead of PND 11. Although immersion fixation is acceptable for the brains of pups at PND 11, brains of pups evaluated at PND 21 should be fixed by *in situ* perfusion (USEPA, 1999).

C. Dose Selection and Dosing

1. Dose Selection

A concurrent control group is required. This group should be a sham-treated group or, if a vehicle is used in administering the test substance, a vehicle control group. An appropriate vehicle should not mask effects and should neither be developmentally toxic nor have effects on reproduction. Animals in the control group should be handled in an identical manner to test group animals (USEPA 1998a).

At least three dose levels of the test article plus a control group should be used. If the test substance has been shown to be developmentally toxic, the highest dose level should be the maximum dose which will not induce *in utero* or neonatal death or malformation sufficient to preclude a meaningful evaluation of neurotoxicity. If a standard developmental toxicity study has not been conducted, the highest dose level, unless limited by the physicochemical nature or biological properties of the substance, should induce some overt maternal toxicity, but should not result in a reduction in weight gain exceeding 20% during gestation and lactation (USEPA 1998a).

The lowest dose should not produce any grossly observable evidence of either maternal or developmental neurotoxicity. The intermediate dose(s) should be equally spaced between the highest and lowest doses used (USEPA 1998a).

2. Dosing

a. Administration of the test article

The route of administration should be determined by potential human exposure. For pesticides this is generally by the oral route. Oral administration is either dietary or by daily gavage (to dams or pups) based on the most recent body weight determination. Other routes of administration may be acceptable, on a case-by-case basis, with ample justification/reasoning for this selection (USEPA 1998a).

b. Dosing Period

Dosing to the dams should cover the period from GD 6 through PND 21. The dosing regimen was extended to PND 21 in the aforementioned DCI issued by the Agency on September 10, 1999. Day 0 of gestation is the day on which a vaginal plug and/or sperm are observed. Gavage dosing should not occur on the day of parturition in those animals who have not completely delivered their offspring. (USEPA 1998a).

c. Verification of Exposure to Pups

In general, it is assumed that exposure to the pups will occur through the maternal milk; however, direct dosing of pups should be considered in those cases where there is a lack of evidence of continued post natal exposure to offspring. Evidence of continuous exposure can be obtained from milk-transfer, pharmacokinetic or milk transfer data. Alternatively, changes in bio markers or offspring toxicity may be used as indications of pup exposure. It is critically important that exposure of the test material to pups be demonstrated for proper validation of the study as well characterization of the dose-response relationship.

The most reliable means to define postnatal exposure is to directly dose the pups by gavage for some or all of early life (Chapin et al., 1997; Beyrouthy et al., 2001; Moser et al., 2001). Direct gavage dosing of pups can be initiated as early as PND 1 or as late as PND 11; however, initiation between PND 4 and 7 is preferred. Initiation of gavage dosing in older pups may result in a relatively long period of minimal exposure for the first week or more of lactation. Without direct dosing, there is concern about the extent to which continuing to dose the dams during lactation provides exposure to the offspring during this dynamic phase of neurological development.

In studies where the dosage to the dam is by gavage, and during dietary studies prior to the onset of diet consumption by pups, postnatal exposure to pups depends exclusively on exposure through the milk. Available data indicate that exposure via milk is variable, both among compounds and across time for a given compound (Dorman et al., 2001).

In dietary studies, postnatal exposure to pups may include some ingestion of test substance by pups during late lactation, but limited data available to EPA suggest that this may only happen between days 18-21 (Gerrish et al., 1998; Hanley & Watanabe, 1985).

The changing nature of actual pup exposure due to growth and metabolism during this critical period brings into question the adequacy of dosing for pups and the impact of that dosing. Methods to determine adequacy of exposure to pups include the following:

- Measurement of test substance in the milk
- Measurement of test substance in tissue (parent and/or major metabolites)
- Measurement of test substance (parent and/or major metabolites) in the plasma of the pups
- Measurement of biomarkers (e.g., cholinesterase inhibition) in pups

D. Maternal Observations

1. General

Daily cage-side observations should be made prior to daily dosing for overt clinical signs of toxicity. The time of onset, degree, and duration of clinical signs should be recorded. Body weight should be measured weekly during gestation and on PND 0, 1, and 21. Food consumption must be measured if the test article is administered in the diet. Water consumption must be measured if the chemical is administered in drinking water (USEPA 1998a).

2. Field Observation Battery (FOB)

Ten dams per group should be observed outside the home cage at least twice during the gestational dosing period (GDs 6-21) and twice during the lactational dosing period (PND 1- 21) for signs of toxicity. Observations should include, but not be limited to, those listed in Table 1. The animals should be observed by trained technicians who are unaware of treatment, using standardized procedures to maximize inter-observer reliability. Where possible, it is advisable that the same observer be used to evaluate the animals throughout a given study. If this is not possible, some demonstration of inter-observer reliability is required. In addition, the description of the FOB procedures used should include where the testing was done, when testing was done with respect to time of dose administration, environmental conditions, the scoring criteria, and duration of the observation period (USEPA 1998a).

Table 1. Functional Observation Battery for Maternal Animals

X	Signs of autonomic function, including: 1) Ranking of degree of lacrimation and salivation, with range of severity scores from none to severe 2) Presence or absence of piloerection and exophthalmus, 3) Ranking or count of urination and defecation, including polyuria and diarrhea 4) Pupillary function such as constriction of the pupil in response to light, or a measure of pupil size 5) Degree of palpebral closure, e.g., ptosis.
X	Description, incidence, and severity of any convulsions, tremors, or abnormal movements.
X	Description and incidence of posture and gait abnormalities.
X	Description and incidence of any unusual or abnormal behaviors, excessive or repetitive actions (stereotypies), emaciation, dehydration, hypotonia or hypertonia, altered fur appearance, red or crusty deposits around the eyes, nose, or mouth, and any other observations that may facilitate interpretation of the data.

E. Litter Observations

The day of completion of parturition (i.e., day of delivery) is designated as PND 0. Daily cage-dose observations should be made for mortality or morbidity and clinical signs of toxicity. The time of onset, degree, and duration of clinical signs should be recorded (USEPA 1998a).

F. Offspring Observations

1. Physical Development

Live pups should be counted, sexed, and weighed individually on PNDs 0, 4, 11, 17, and 21 and at least once every two weeks thereafter. The age of vaginal opening and preputial separation should be determined and the body weight at attainment recorded (USEPA 1998a). General procedures for these determinations may be found in Adams et al. (1985) and Korenbrot et al., (1977). Other developmental landmarks such as pinna unfolding, incisor eruption, etc., are not specified in the DNT guideline, but may be included in the study.

2. Post Weaning Observations

After weaning on PND 21, offspring should be examined twice daily for mortality and moribundity and once daily for clinical signs of toxicity. The time of onset, degree, and duration of clinical signs should be recorded. Individual body weight data are recorded at least every two weeks (USEPA 1998a).

3. Neurobehavioral Evaluations

A detailed description and information of the types of equipment and automated devices used, procedures employed for calibrating and ensuring the equivalence of devices, and the balancing of treatment groups in testing procedures used in neurobehavioral evaluations should be included in the report.

a. FOB

A total of 10 male and 10 female offspring per dose group should be examined outside the home cage on PNDs 4, 11, 21, 35, 45, and 60. The same animals are to be used at all time points. The offspring should be observed by trained technicians who are unaware of treatment, using standardized procedures to maximize inter-observer reliability. Where possible, it is advisable that the same observer be used to evaluate the animals throughout a given study. If this is not possible, some demonstration of inter-observer reliability is required. Description of the observation procedures, including scoring criteria (may vary with age of animals), should be available in the study report.

Assessments should be age-appropriate for the developmental stage being observed. At a minimum, the same endpoints outlined above for the maternal FOB (Table 1) should be monitored as appropriate for the developmental stage being observed (USEPA 1998a). On postnatal days 4 and 11, the animals can be tested for surface righting reflex then observed in the open field for approximately one minute. In the open field, PND 4 pups should be assessed for activity and any physical abnormalities in appearance or gait while PND 11 pups should be assessed for activity, number of rearings, grooming, urination, and physical abnormalities. Methods on days 21, 35, 45, and 60 should be similar to the procedures used for the dams.

b. Motor Activity

Motor activity should be monitored on at least 10 male and 10 female offspring per group specifically on PND 13, 17, 21, and 60± 2. The same animals are to be used at all time points. Motor activity should be monitored by an automated activity recording apparatus. The device should be capable of detecting both increases and decreases in activity. Each device should be

tested by standard procedures to ensure, to the extent possible, reliability of operation across devices and across days for any one device. In addition, treatment groups must be balanced across devices. Each animal should be tested individually. The test session should be long enough for motor activity to approach asymptotic levels by the last 20% of the session for nontreated control animals. All sessions should have the same duration. Treatment groups should be counter-balanced across test times. Activity counts should be collected in equal time periods of no greater than 10 minutes duration. Efforts should be made to ensure that variations in the test conditions are minimal and are not systematically related to treatment. Among the variables that can affect motor activity are sound level, size and shape of the test cage, temperature, relative humidity, light conditions, odors, use of home cage or novel test cage, test order, and environmental distractions (USEPA 1998a). Additional information on the conduct of a motor activity study may be obtained in the *Health Effects Test Guidelines OPPTS 870.6200 Neurotoxicity Screening Battery*.

c. Auditory Startle Test

An auditory startle habituation test should be performed on at least 10 male and 10 female offspring per group around the time of weaning and around PND 60. Day of testing should be counter-balanced across treated and control groups. While use of pre-pulse inhibition is not a requirement, it is strongly recommended. In performing the auditory startle task, the mean response amplitude on each block of 10 trials (5 blocks of 10 trials per session on each day of testing) should be made (USEPA 1998a). Further details on the conduct of this test may be obtained in Ison (1984).

d. Learning and Memory Tests

A test of associative learning and memory should be conducted on at least 10 male and 10 female offspring per group around the time of weaning and around PND 60. Either different animals should be used at each time point, or the same animals can be used with different tests at each age. Since all animals can not logistically be tested on the same day, each day of testing should be counter-balanced across treated and control groups. Some flexibility is allowed in the choice of tests for learning and memory in weanling and adult rats such that the same or separate tests may be used at these two stages of development. However, the tests must be designed to fulfill the following two criteria:

- First, learning must be assessed either as a change across several repeated learning trials or sessions, or in tests involving a single trial, with reference to a condition that controls for no associative effects of the training experience.

- Second, the tests should include some measure of memory (short-term or long-term) in addition to original learning (acquisition). If the tests of learning and memory reveal an effect of the test compound, it may be in the best interest of the sponsor to conduct additional tests to rule out alternative interpretations based on alteration in sensory, motivational, and/or motor capacities.

In addition to the above criteria, it is recommended that the test of learning and memory be chosen on the basis of its demonstrated sensitivity to the class of compound under investigation, if such information is available in the literature (USEPA 1998a). In the absence of such information, examples of tests that could be made to meet the above criteria include: delayed-matching-to-position, as described for the adult rat (Bushnell 1988) and for the infant rat (Green and Stanton 1989); olfactory conditioning (Kucharski and Spear 1984); and acquisition and retention of schedule-controlled behavior (Campbell and Haroutunian 1981; Cory-Slechta et al. 1983). Additional tests for weanling rats (Spear and Campbell 1979; Krasnegor et al. 1987) and for adult rats (Miller and Eckerman 1986) have been described elsewhere.

A number of water maze tasks have been developed to assess learning and memory. In general, all water maze tests rely on escape from water as the primary motivation and are therefore aversively-motivated tasks. The water maze tests include: **Morris** water maze, **Biel** water maze, **Cincinnati multiple-T** water maze, as well as the single choice, positional habit **E- and M-shaped mazes**. Depending on the type of test used, the specific parameters used will vary. Appropriate data may include trials to criterion, number of errors, percent correct, latency of response, number of quadrants entered, or others.

The **Morris water maze** is a water-filled swimming maze consisting of a circular pool constructed of white plastic (90 cm diameter, 30 cm deep at weaning, and 140 cm diameter, 45 cm deep at day 60). The maze is filled with water approximately $29\pm 3^{\circ}\text{C}$ made opaque with a non-toxic opacifier. A platform 6 cm square for weanling rats and 10 cm square for adult rats is located at a fixed point in the pool, concealed approximately 1.5 cm below the surface of the water. Three starting points are identified at the perimeter of the pool and a number of visual clues placed on the walls of the pool and outside the pool are available to assist learning. A series of 3 trials is conducted on each of 4 consecutive days. On the first trial the rat is placed on the escape platform for 30 seconds prior to testing. The animal is then placed into the water at the perimeter of the pool and allowed a maximum of 90 seconds to swim to the platform. A different starting point is used for each trial. The time to reach the platform and the number of quadrants of the pool crossed are recorded.

The **Biel water maze** uses a water-filled, six unit T-maze. Each assessment evaluates swimming ability on day 1 followed by maze learning ability on days 2-5. On days 2 and 3 each animal is tested on the forward direction through the maze and on days 4 and 5 each animal is tested in the reverse direction through the maze. After a three-day rest period, each animal is tested for memory recall of both directions through the maze. The mean number of errors (all four feet into an incorrect channel) and mean escape time for each test day were considered measures of maze-learning ability.

The **M-water maze** is made of opaque Plexiglas with 5-inch wide corridors. The walls are 16 inches high with approximately 7.5 inches of water at $22\pm 1^{\circ}\text{C}$. For each test trial, the rat is placed at the base of the M-maze stem, between the two lateral arms. On the learning trial (first trial), the rat is required to enter both arms of the maze before being provided access to the exit ramp to escape the water. The initial arm chosen on the learning trial is designated the incorrect goal during the subsequent trials (15 maximum). Rats failing to make a correct goal choice within 60-seconds in any given trial are led to the correct goal with the exit ramp and then removed from the water. The inter-trial interval is approximately 15 seconds. Each rat is required to reach a criterion of 5 consecutive error-free trials to stop the test session. Retention is assessed by testing the same animal 7 days later. Latency (in seconds) to choose the correct goal or the maximum 60-second interval and the number of errors (incorrect turns) during each trial are recorded.

Testing for **passive avoidance conditioning** is done in individual isolation cubicles each with a single shuttle cage. Each shuttle cage (approximately 7 x 7 inches) is separated into two equal-sized compartments by a centrally-located sliding door. The two compartments are identical except that the walls in one compartment are lined with black film (dark side) and the walls in the other compartment are not lined and this compartment is illuminated with a high-intensity lamp. The lamp is switched on at the beginning of each trial and remains on until the rat crosses into the dark compartment or the trial ends. The cage floor is constructed of a stainless steel grid and the movement of the rat from the light to dark side is detected by a photocell. Rats are placed individually into the shuttle cage facing toward the light. After 20 seconds, the light is switched on and the door separating the compartments is opened. When the rat crosses into the dark side, the door is closed, and a brief, mild shock (0.5 sec, 0.5mA) is delivered. The procedure is repeated until the rat either remains in the bright side for 180 seconds for two consecutive trials or until 15 trials have elapsed (whichever occurs first). Retention is assessed by testing the animals again after a 7-day interval.

4. Postmortem Evaluations

a. Brain Weight and Fixation

Neuropathological evaluation should be conducted on animals on PND 11 and at the termination of the study. Alternatively, instead of PND 11 it is acceptable to conduct evaluations on PND 21 if dosing of the dams was extended to day 21.

At PND 11 one male or one female pup should be removed from each litter such that equal numbers of male and female offspring are removed from all litters combined. The pups should be killed humanely and immediately thereafter the brains should be removed and weighed. **Of these animals, brains from 10 male and 10 female pups per dose should be immersion-fixed in an appropriate aldehyde fixative for neuropathological analysis.** After fixation, paraffin embedding is acceptable but plastic embedding is preferred and recommended. Tissue blocks and slides should be appropriately identified when stored. Histological sections should be stained with hematoxylin and eosin, or a similar stain according to standard published protocols (Bennett et al. 1976; Ralis et al. 1973; Luna 1968).

At PND 21 and study termination, one male and one female from each litter should be killed humanely and immediately thereafter the brain should be removed and weighed. In addition, **10 animals per 10 per dose group (one male or one female per litter) should be sacrificed at the termination of the study for neuropathological analysis.** These animals should be anesthetized and tissues fixed by perfusion. Central and peripheral nervous tissues should be dissected, preserved in paraffin (CNS tissues) or plastic (PNS tissues), blocked, sectioned, and stained with hematoxylin and eosin. Details of these procedures for adult animals have been published in the *Health Effects Test Guidelines OPPTS 870.6200 Neurotoxicity Screening Battery*.

b. Neuropathology

i. Qualitative Analysis

The purposes of the qualitative examination are to identify regions within the nervous system exhibiting evidence of neuropathological alterations, to identify types of neuropathological alteration resulting from exposure to the test substance, and to determine the range of severity of the neuropathological alteration. Samples from all major brain regions (olfactory bulbs, cerebral cortex, hippocampus, basal ganglia, thalamus, hypothalamus, midbrain [tectum, tegmentum, cerebral peduncles], brain stem, and cerebellum) of pups and adults and peripheral nervous tissues from adults must be examined. Representative histological sections from the tissue samples should be examined microscopically by an appropriately trained pathologist for evidence of neuropathological alterations.

The following stepwise procedure is recommended for the qualitative analysis. First, sections from the high dose group are compared with those of the control group. If no evidence of neuropathological alteration is found in animals of the high dose group, no further analysis is required. If evidence of neuropathological alterations is found in the high dose group, then animals from the intermediate and low dose group are examined. Subject to professional judgment and the kind of neuropathological alterations observed, it is recommended that additional methods such as Bodian's or Bielchowsky's silver methods and/or immunohistochemistry for glial fibrillary acid protein be used in conjunction with more standard stains to determine the lowest dose level at which neuropathological alterations are observed (USEPA 1998a).

However, Agency experience, based on the DNT studies reviewed to date, suggest that it may be preferable to evaluate all treatment groups at the same time, to allow for interpretation of neuropathology data in the context of the contemporaneous analysis of all the data from the other groups. It also avoids delays in study reviews that could result if questions arise about the need for lower dose evaluations. At a minimum, tissues from all dose groups should be taken to the block stage.

In addition to more typical kinds of cellular alterations (e.g., neuronal vacuolation, degeneration, necrosis) and tissue changes (e.g., astrocytic proliferation, leukocytic infiltration, cystic formation) particular emphasis should be paid to structural changes indicative of developmental insult including but not restricted to:

- Gross changes in the size or shape of brain regions such as alterations in the size of the cerebral hemispheres or the normal pattern of foliation of the cerebellum.

- The death of neuronal precursors, abnormal proliferation, or abnormal migration, as indicated by pyknotic cells or ectopic neurons, or gross alteration in regions with active proliferative and migratory zones, alterations in transient developmental structures (e.g., the external germinal zone of the cerebellum).
- Abnormal differentiation. While more apparent with special stains, this may also be indicated by shrunken and malformed cell bodies.
- Evidence of hydrocephalus, in particular enlargement of the ventricles, stenosis of the cerebral aqueduct and general thinning of the cerebral hemispheres (USEPA 1998a).

Further guidance for neuropathological examination for indication of developmental insult to the brain can be found in the following articles: (Friede 1975; Suzuki, 1980; DeGroot et al., 2005a, 2005, b).

ii. Subjective Analysis

If any evidence of neuropathological alterations is found in the qualitative examination, then a subjective diagnosis should be performed for the purpose of evaluating dose-response relationships. All regions of the brain exhibiting any evidence of neuropathological changes should be included in this analysis. Sections of each region from all dose groups will be coded as to treatment and examined in randomized order. The frequency of each type and the severity of each lesion will be recorded. After all sections from all dose groups including all regions have been rated, the code will be broken and statistical analyses performed to evaluate dose-response relationships. For each type of dose related lesion observed, examples of different ranges of severity should be described. The examples will serve to illustrate a rating scale such as 1+, 2+, and 3+ for the degree of severity ranging from very slight to very extensive (USEPA 1998a).

iii. Simple Morphometric Analysis

Brain tissue collected for morphometric analyses should be embedded in appropriate media at all dose levels at the same time in order to avoid shrinkage artifacts that may be associated with prolonged storage in fixative.

Since disruption of developmental processes is sometimes more clearly reflected in the rate or extent of growth of particular brain regions, some form of morphometric analysis is required to be performed on PND 11 (or PND 21) and at the termination of the study to assess the structural development of the brain. At a minimum, this would consist of a reliable estimate of the thickness of major layers at representative locations within the neocortex, hippocampus, and cerebellum (Rodier and Gramann, 1979;USEPA 1998a).

Examples include:

- Frontal cortex thickness (dorsal portion of the cerebral cortex within the coronal section passing through the region of the optic chiasm)
- Parietal cortex thickness (dorsolateral portion of the cerebral cortex within the coronal section taken through the optic chiasm)
- Caudate putamen horizontal width (coronal section taken at the level of the optic chiasm)

- Corpus callosum (thickness at the midline)
- Hippocampal gyrus (greatest dorsal-ventral thickness)
- Cerebellum (roof of the fourth ventricle to the dorsal surface)

G. Biomarkers of Exposure

With some chemicals or chemical groups, measurement of specific biomarkers may provide useful information. If analysis of biomarkers is considered necessary, the protocol details should be worked out with the agency prior to study initiation.

For example, cholinesterase activity measurements are required if the test article is an organophosphate or carbamate insecticide. Blood and brain samples should be collected at sacrifice from dams on PND 21 and from offspring on PND 4, 21, and 60. It should not have been necessary to pool samples from pups in order to attain a sufficient volume of tissue for cholinesterase measurement. However, if PND 4 blood samples needed to be pooled, samples should only have been combined from fetuses of the same sex within the same litter, not from fetuses of different litters.

H. Statistical Evaluations

Statistical analyses of maternal and offspring data should be conducted with careful considerations of study design, sample size, the endpoint evaluated, the variability in the incidence of the endpoint, effect of gender, time of measurement, the robustness of the data, and the influence of litter on analytical outcome. For the selection of appropriate statistical methods and data analysis, consultation with statistician familiar with appropriate analyses of data is recommended.

IV. STUDY INTERPRETATION

A. Endpoints of Maternal Toxicity

The toxicity endpoints measured in the dams in the DNT are few and rather crude (e.g., survival, body weights, feed consumption, clinical observations, and optional organ weights) when compared to the large number of endpoints measured in the pups in this study. Therefore, it would be helpful to include comparison of the observations from developmental and reproductive toxicity studies and other toxicity tests when interpreting maternal toxicity. .

In general, maternal toxicity is characterized as at or less than 10% mortality, at or less than 10% reduction in body weight gain, clinical signs of toxicity, and/or evidence of dose limiting toxicity to a target organ.

A brief overview for the evaluation of maternal toxicity endpoints commonly encountered during gestation and lactation periods is presented below. For a comprehensive guidance on the evaluation/analyses of these endpoints, the reader is advised to consult the following documents:

- SEP for Teratology Studies (Hazard Evaluation Division, 1985),
- SEP for Developmental Toxicity Studies (Health Effects Division, 1993),
- SEP for Reproductive Toxicity Studies (Health Effects Division, 1993),
- Guideline for Developmental Toxicity Risk Assessment, USEPA 1991,
- Guideline for Reproductive Toxicity Risk Assessment, USEPA 1996
- Guidance Document for Neurotoxicity Testing (OECD, 2004); and
- Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment (OECD, 2008)

1. Mortality and Clinical Signs

Neurotoxicity, developmental, reproductive, and range-finding toxicity studies conducted prior to the DNT study should provide information on the clinical signs of toxicity characteristic of the test substance in adult animals. On a case-by-case basis these studies may be used to support equivocal evidence of maternal toxicity and some reference to these studies must be given in the DER. Special emphasis should be placed on the detailed clinical observations conducted in the dams as described Section III.D.2.

The DNT study protocol considered in this SEP does not fully evaluate neurotoxicity in the dams. Therefore, a dose-response relationship should be evaluated before clinical signs and mortality are attributed to administration of the test compound. Where a dose-relationship is not clear, the effects of disease, environmental conditions, or technical errors should be eliminated as possible causes of the observed mortality and clinical signs. These factors must be considered along with the background information from the preliminary studies to completely evaluate the results.

Consequently, maternal effects due to causes other than test substance toxicity should always be considered when evaluating the adequacy of the study. Maternal observations such as organ weights and histopathology are not included, and the highest dose group may be the only group with increased clinical signs or mortality. In this case, comparisons of the signs reported in the DNT study with signs observed in other studies in the same species, should be used to establish the observations as compound-related. Presence or absence of statistical significance does not constitute or refute a difference as being treatment-related; biological significance, designated trends, data from other studies etc are also important considerations.

2. Maternal Body Weight

Absolute body weight and body weight changes can be sensitive indicators of systemic toxicity and are often used as the basis for the determination of NOAEL/LOAEL for maternal toxicity. Animals must be randomly assigned to groups based on evenly distributing females by body weight. If body weight is the criterion, this assignment evenly distributes the variation in the initial body weight among the groups. When this assignment is by a specified randomization procedure (e.g., blocked randomization), initial mean body weights and their variances become more homogeneous. Theoretically, the body weight gain (or percentage change in body weight) of all the treated groups should be comparable to that of the controls during the period prior to treatment. If they are not, the predosing differences should be accounted for in an analysis of subsequently observed weight gain differences. Likewise, body weight changes may provide more information than a daily body weight measured during treatment

In dietary studies, it is possible that body weight changes may result from reduced palatability of the diet due to the presence of the test article rather than an effect of the chemical. A reduction in body weight may be due to direct effect of the test material on the metabolism or digestive process of an organism or may result from a decrease in food intake unrelated to the inherent toxicity of test material (palatability or decreased appetite). Reduced palatability is suggested by decreased food consumption in the absence of other indications of toxicity. Therefore, body weight data should be assessed along with food consumption data for the calculation of food efficiency. Maternal body weight gain can also be influenced by changes in fetal growth. Comparisons of the mean maternal weight prior to and after delivery may also serve as an indicator of maternal toxicity. An alternate but less desirable estimate of maternal weight change during gestation can be obtained by subtracting the litter weight from the maternal weight gain.

3. Food and water consumption

Maternal food and water consumption data are not required for DNT studies using gavage administration. In those studies using dietary or drinking water routes for administration of the test material, food and water consumption are used to calculate the dose to the animal. However, food and water consumption data may be useful for assessment of maternal effects regardless of the route of administration since decreases in consumption may lead to decreases in body weight or indicate decreased palatability of the test diet in dietary studies. Small changes in water consumption, even if statistically significant, are difficult to evaluate, but if such decreases are dose- or compound-related they may indicate effects on excretory function, appetite, or water consumption.

Determining food efficiency when food consumption is affected by the test article may be helpful in characterizing maternal toxicity, but these results are subject to the variability associated

with the food consumption and body weight components of the calculation. Efficiency of food utilization (E_f) can be calculated as follows:

$$E_f = (\text{g body wt change per unit time}) / (\text{g food consumption per unit time}) \times 100$$

This calculation gives the percentage efficiency with which the animal converts food for maintenance. Low efficiency compared with control indicates toxicity in the consuming animals.

A complicating factor in the interpretation of food and water consumption data is the amount attributable to the pups during lactation. By approximately PND 10 the pups will begin to play in the food or with the water bottle resulting in significant spillage or wastage. By late lactation the pups will be eating and drinking on their own in addition to nursing. These actions by the pups can result in not only over estimation of consumption by the dam, but in significant direct exposure of the pups to the test article. Selected group mean body weights and food consumption values for pregnant or nursing dams should be presented as follows:

Table 2. Selected mean (\pmSD) Maternal Body weight, Body Weight Gain and Food Consumption				
Observations/study interval	Dietary concentration (ppm)¹			
	Control	LDT	MDT	HDT
Gestation (n=)				
Body wt. Gestation day 6 (g)				
Body wt. Gestation day 13 (g)				
Body wt. Gestation day 20 (g)				
Wt. gain gestation days 6-20 (g)				
Food consumption gestation days 6-13 (g/day)				
Food consumption gestation days 13-20 (g/day)				
Lactation (n=)				
Body wt. lactation day 0 (g)				
Body wt. lactation day 4 (g)				
Body wt. lactation day 7 (g)				
Body wt. lactation day 14 (g)				
Body wt. lactation day 21 (g)				
Wt gain lactation days 0-21(g)				
Food consumption lactation days 0-7 (g/day)				
Food consumption lactation days 7-14 (g/day)				
Food consumption lactation days 14-21 (g/day)				

¹ = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

4. Reproductive Performance

Due to the design of the DNT study, the data available to assess reproductive performance are limited to gestation length, and effects on parturition and live litter production. The length of gestation is the interval from positive evidence of copulation to parturition. It can be calculated from the individual breeding records of the animals and is summarized as a mean value. Significant decreases in gestation length may result in decreased offspring birth weight and survival. Significant increases may result in dystocia (difficult labor and delivery) and ultimately in death or physical impairment of the dam and/or offspring. In addition, if lengthened gestation is not complicated by dystocia, the offspring may be larger and heavier at birth, which could, without the application of more sophisticated statistical analysis techniques, mask treatment-related differences in body weight gains of the offspring later during lactation.

Decreases in mating or fertility may be important indicators of maternal health and the general status and suitability of the strain for use in toxicity testing. Changes in the gestation length may indicate effects on parturition and/or hormone homeostasis, while decreases in postnatal viability may reflect maternal and/or developmental toxicity. Results for the maternal animals should be presented as shown in table below:

Table 3. Reproductive Performance				
Observation	Dietary concentration (ppm)¹			
	Control	LDT	MDT	HDT
Number mated				
Number (%) pregnant				
Number Delivered				
Gestation Index (%)				
Gestation Length (days)				
Mean Number (\pm SE) of Implantation Sites				
Number (%) with Stillborn pups				
Number with Complete Litter Loss				

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

¹ = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

5. Organ Weight and Pathology

Maternal organ weight and pathology data are not required in DNT studies. However, such data may be included and can be used in support of maternal toxicity, especially where the test substance causes little or equivocal body weight decreases but is associated with target organ toxicity.

B. Endpoints of Offspring Toxicity

A brief overview is presented below for the evaluation of offspring toxicity characterized as viability, clinical signs, body weight/body weight gains, and developmental landmarks. For a comprehensive guidance on the evaluation/analyses of these endpoints, the reviewers are advised to consult the following SEP:

- SEP for Reproductive Toxicity Studies (Health Effects Division, 1993)
- Guidance Document on Mammalian Reproductive Toxicity Testing (OECD, 2008)

1. Viability and Clinical Signs

A decrease in live litter size may be indicative of a reproductive effect (resulting from changes in a number of maternal parameters) or an effect on fetuses. However, most of the maternal endpoints (e.g., ovulation, fertilization, and implantation) occurred prior to start of dosing and alterations suggestive of effects on these endpoints should be considered carefully. Mean litter size is often affected by cannibalism requiring close monitoring of delivering dams by the laboratory technical staff and removal of dead pups from the nesting cage to prevent loss of information. Maternal stresses such as disease, environmental conditions, etc (see Section IV.A.1) can also affect live birth outcome, survival, and behavior.

At birth, all pups should be examined for external anomalies as well as for viability. The number of viable, stillborn, and cannibalized members of each litter should be recorded. Throughout lactation, both maternal and pup effects can influence offspring viability as shown in table below.

Table 4. Factors influencing offspring viability
Developmental effects on the young (abnormal and/or inadequate organ development or change in offspring behavior) as a result of <i>in utero</i> exposure may be evident at any time point.
Varied effects of maternal toxicity: -maternal neglect (behavioral change) -inadequate milk production (e.g., endocrine change, inadequate nutrition)
Postnatal toxicity due to the presence of the agent in the milk.

The following offspring survival indices can aid in determining when decreases in survival are occurring during lactation:

$$\text{Live birth index (\%)} = (\text{No. pups born alive} / \text{No. pups born}) \times 100$$

$$\text{Viability index (\%)} = (\text{No. pups alive PND 4 – pre culling} / \text{No. pups born alive}) \times 100$$

$$\text{Lactation index (\%)} = (\text{No. pups alive PND 21} / \text{No. pups alive PND 4}) \times 100$$

Clinical signs of toxicity in the offspring are important, both alone and in context of other data. Careful recording of onset and duration of clinical signs observed during lactation allows correlation with survival and behavior. In addition, observations made from cageside examination can be related to observations made during the FOB and should be followed into postweaning.

Litter size and viability (*survival*) results from pups during lactation should be presented as shown in Table below.

Table 5. Litter Size and Viability				
Observation	Dietary concentration (ppm)¹			
	Control	LDT	MDT	HDT
Total number born				
Number born live				
Number born dead				
Sex Ratio Day 0 (% %)				
# Deaths Days 0-4 (%)				
# Deaths Days 4-21 (%)				
Mean litter size:				
Day 0				
Day 4 (before culling)				
Day 4 (after culling)				
Day 11				
Day 17				
Day 21				
Live birth index				
Viability index				
Lactation index				

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

¹ = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

2. Physical Development

a. Body Weight

In addition to the offspring indices described in Section B.1, a very important indication of offspring health status is the weight of the surviving pups. Absolute body weight and body weight changes can be sensitive indicators of systemic toxicity and can be used as the basis for the determination of NOAEL/LOAEL for offspring toxicity. Live pups should be counted and litters weighed, by weighing each individual pup (optional) at birth, or soon thereafter, and on days 4, 7, 14, and 21 after birth.

A change in offspring body weight is a sensitive indicator of developmental toxicity, in part because it is a continuous variable. In some cases, offspring weight reduction may be the only indicator of developmental toxicity. While there is always a question as to whether weight reduction is a permanent or transitory effect, little is known about the long-term consequences of short-term fetal or neonatal weight changes.

Pup body weight data should be evaluated concurrently with pup survivability data. As discussed earlier, an increase in offspring mortality after postnatal day 4 may result from nutritional deficiency, maternal neglect (behavioral change), or directly from the toxicity of the chemical tested (concentration in the milk due to the lipophilic nature of the agent or ingested in feed). Therefore, an increase in offspring mortality without impairment of pup body weights may, at least, rule out the possibility of nutritional deficiency; whereas concurrent increases in offspring mortality and decreases in pup body weights may result from any or all of these factors.

Regardless of the exact etiology of increased mortality and decreased pup body weights, these findings are generally considered to be toxic effects. Several other factors should be considered in the evaluation of fetal or neonatal weight changes; for example in polytocous animals, fetal and neonatal weights are usually inversely correlated with litter size, and the upper end of the dose-response curve may be affected by smaller litters and increased fetal or neonatal weight. Additionally, alterations in pup body weight can also be the result of an *in utero* effect manifested postnatally.

It is also important to keep in mind that litter size has an important influence on pup weight. Mean pup weight shows a slight but consistent decrease with increasing litter size when litters number six pups or greater (Khera et al., 1989). If litter size is greater in treated groups than in the control group, whether due to chance or as a compound-related effect, decreased pup weights may be expected.

The weight of the pups at weaning (lactation day 21) is another important parameter that should be considered in the evaluation process. A difference in neonatal birth weight between control and treated pups does not necessarily imply that a difference in weaning weight will ensue. The weaning weight may be similar to controls, suggesting a reversible effect, or remain altered, suggesting an irreversible effect. However, attainment of expected weight at weaning does not demonstrate that untested functional effects have neither occurred nor persisted.

In late lactation, the pups become very active and begin to play in and eat the food presented to the dam. For treated groups, this results in additional exposure of the pups to the test substance. They may be receiving it in the milk from nursing, in the food consumed, and even possibly dermally from playing in the food container. This additional exposure to the pups may result in evidence of increased toxicity late in lactation, including treatment-related decreased in body weight, mortality, and adverse clinical findings.

In general, the concurrent control data should be the basis for examining the data of treated groups for compound-related effects. In some incidences, if the concurrent control data raises questions or concerns, then the recent historical control data from the testing laboratory (in combination with the concurrent control data) could be used in aiding interpretation of unusual or inconsistent findings in pup body weight data.

For proper evaluation/analyses, selected mean pre-weaning and post-weaning pup body weight data should be presented as shown in the tables below:

Table 6. Mean (\pm SD) Pre-weaning Pup Body Weights (g)

Postnatal day	Dietary concentration (ppm) ¹							
	Control	LDT	MDT	HDT	Control	LDT	MDT	HDT
	Males				Females			
1								
4 (before culling)								
4 (after culling)								
11								
17								
21								

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

1 = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

Table 7. Mean (\pm SD) Post-Weaning Pup Body Weights (g)

Postnatal day	Dietary concentration (ppm) ¹							
	Control	LDT	MDT	HDT	Control	LDT	MDT	HDT
	Males				Females			
35								
49								
[#]								
[#]								

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

1 = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

b. Developmental Landmarks

The evaluation of developmental parameters in the offspring can provide useful information regarding sexual maturation as well as indications of hormonally-mediated effects. The onset of sexual maturity (puberty), generally a body weight-dependent milestone, is evaluated by examination of females for opening of the vaginal orifice and by determination of preputial separation in the males (Adams et al., 1985; Korenbrot et al. 1977). Typically, these events occur at approximately PND 40-45 in male rats and PND 30-35 in female rats, although there may be some strain-related differences (OECD, 2008). Other developmental milestones, such as measurements (time-to-event) of eye opening, pinna unfolding, incisor eruption, piliation, righting reflex, cliff avoidance, and negative geotaxis, may be included as part of the clinical observations or as part of the age-appropriate FOB. Delays in, absence of, or changes in any of these normal processes, whether or not linked to body weight, may indicate a toxic response to treatment. Summary of any biologically relevant effects on vaginal opening and preputial separation or other developmental milestones should be presented as shown below:

Table 8. Mean (\pm SD) Age of Sexual Maturation (days) and Body Weight At Time of Attainment

Parameter	Dietary concentration (ppm) ¹			
	Control	LDT	MDT	HDT
N (M/F)				
Preputial separation (males)				
Vaginal opening (females)				
Body weight (g)				

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

1 = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

3. Neurobehavioral Evaluations

Behavior reflects the integration of the various functional components of the nervous system. Changes in behavior can arise from a direct effect of a toxicant on the nervous system, or indirectly from its effects on other physiological systems. The presence of systemic toxicity may complicate, but does not preclude, interpretation of behavioral changes as evidence of neurotoxicity. During the evaluation of the study, data from each of the behavioral tests are assessed separately and in isolation of each parameter. However, determination of whether there is an indication of a treatment-related effect should rely on an integrative evaluation of all the parameters in the study.

Data with large variability should be examined closely. Data for the control group should give an indication of the quality of the data set. If the results of the control animals appear reasonable, then the data for treated animals can be examined for evidence of treatment-related effects. On the other hand, a high degree of variability in the control group may indicate problems with the ability of the testing facility to properly perform the test or may indicate lack of sensitivity in the test method to accurately measure the endpoint in question. Consequently, the results of the concurrent controls should be evaluated closely prior to examining the results of the treated animals.

A brief overview for the evaluation of functional observations, motor activity, auditory startle response, and learning and memory is presented below. For a comprehensive guidance on the evaluation/analyses of these endpoints, the reviewers are advised to consult the following journal articles:

- Tyl, R.W, et al. 2008. Identification and interpretation of developmental neurotoxicity effects. A report from the ILSI Research Foundation/Risk Institute expert working group on neurodevelopmental endpoints. *Neurotoxicology and Teratology*. 30: 349-381.
- Raffaele, K.C. et al. 2008. Determining normal variability in a developmental neurotoxicity test. A report from the neurodevelopmental endpoints. *Neurotoxicology and Teratology*. 30: 288-325
- Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment (OECD, 20008)

a. Functional Observation Battery (FOB)

An FOB is designed to detect and quantify major overt behavioral, physiological, and neurological signs. The test article may alter the maturation or appearance of sensorimotor reflexes such that significant alteration in or delay of reflexes is evidence of a neurotoxic effect (USEPA 1998b). Although the data collected may not allow differentiation between neurotoxicity and systemic toxicity, effects are still of concern. Of importance to the validity of the FOB results are age-appropriate tests which account for ontogeny of behaviors and the use of the same animals at each time point.

Data for the FOB evaluations varies considerably across laboratories. Therefore, it is critical to look for trends across days, as well as differences in treatment groups for any given evaluation. It is also essential to take into account that not all neural systems are fully developed at birth, and therefore, the expected responses may differ for specific ages of animals.

Tests must be appropriate for the developmental stage of the offspring since pre-weaning animals act and respond differently than adults. A change in scoring criteria may be necessary to reflect “normal” in young animals. For example, young animals have a slightly uncoordinated gait and may show a slight tremor when placed in the open field. Scoring criteria for adults would consider these observations abnormal when they might be normal for the age group (Moser 2000). Also, air righting ability does not fully develop until about the time of weaning so surface righting is a more appropriate test in young animals.

Additionally, some tests can not be adequately performed on young rats because the response has not developed or because the data generated are too variable for meaningful interpretation (Moser 2000). Endpoints which may be excluded from the FOB for young rats include pupil response (as eye opening may not have occurred), ataxia score (uncoordinated gait may be normal), landing foot splay (hind limbs do not always support), home cage activity and ease of removal (litters huddle), approach and touch responses (many young rats show no response), and body temperature (young rats can not thermoregulate consistently).

Open field activity, rearing activity, and grip strength show clear age-related responses. Before eye opening, rats show very little exploring or rearing activity. Both open field activity and rears increase from PND 17 to PND 27, decrease at PND 40, and increase again at PND 70 (Moser 2000). Preweaning rats have not developed sufficient grip strength to register a response on the usual strain gauge. However, post-weaning, both fore- and hindlimb grip strengths increase with increasing muscle mass and neuromuscular ability; gender differences are apparent after PND 40 with males greater than females at PND 70 (Moser 2000).

The Guidelines for Neurotoxicity Risk Assessment (USEPA 1998b) state that:

“The relevance of statistically significant test results from an FOB is judged according to the number of signs affected, the dose(s) at which effects are observed, and the nature, severity, and persistence of the effects and their incidence in relation to control animals. In general, if only a few unrelated measures in the FOB are affected, or the effects are unrelated to dose, the results may not be considered evidence of a neurotoxic effect. If several neurological signs are affected, but only at the high dose and in conjunction with other overt signs of toxicity, including systemic toxicity, large decreases in body weight, decreases in body temperature, or debilitation, there is less persuasive evidence of a direct neurotoxic effect. In cases where several related measures in a battery of tests are affected and the effects appear to be dose dependent, the data are considered to be evidence of a neurotoxic effect, especially in the absence of systemic toxicity”.

Any treatment-related findings observed should be summarized as shown in Table 9. Data should be included for all statistically significant findings, and for any findings the reviewer feels are toxicologically relevant (even if not statistically significant [e.g., an incidence of 3/10 for a parameter where controls are 0/10 would warrant inclusion]). If significant effects are found, data from all groups, time points, and both sexes should be included for that parameter (so that the effect can be compared across time and groups). Include severity information if there are changes in severity.

Table 9. Functional Observational Battery Results (Incidence)				
Observation	Dietary concentration (ppm)¹			
	Control	LDT	MDT	HDT
Males				
<u>Type of observation -1</u>				
-PND #				
-PND #				
-PND #				
-PND #				
-PND #				
<u>Type of observation -2</u>				
-PND #				
-PND #				
-PND #				
-PND #				
-PND #				
Females				
<u>Type of observation -1</u>				
-PND #				
-PND #				
-PND #				
-PND #				
-PND #				
<u>Type of observation -2</u>				
-PND #				
-PND #				
-PND #				
-PND #				

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

¹ = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

b. Motor Activity

Motor activity should be monitored in at least 10 male and 10 female offspring per group specifically on PND 13, 17, 21, and 60± 2. Motor activity (MA) is usually quantified as the frequency of movements over a period of time. The total counts generated during a test period will depend on the mechanism, the size and configuration of the testing apparatus. Effects of agents on MA can be expressed as absolute activity counts or as a percentage of control values.

Certain alterations in the nervous system development in the young test animals are often measured by changes in motor activity. Motor activity can be seen as large or small movements. Large movements frequently involve changes in the center of gravity of a test animal, and movement can be horizontal or vertical. Small movements consist of several types; they can be tremors, reflexive movements, or directed movements. The movement pattern for developmental

motor activity in pups varies as the animals grow. Typically for non-treated animals, MA is lowest in 13-day-old pups, before their eyes open and when their motor skills are quite limited. MA increases from PND 13-17 and decreases from PND 17-21. From PND 21 to 60, the MA level of the test pups generally does not show significant fluctuation. Habituation is not seen on PND 13 but is present by PND 21 (Ruppert et al. 1985; Moser, 2000; Cory-Slechta et al. 2001; Raffaele et al, 20087; Tyl et al., 2008).

Motor activity should be evaluated both with regard to total activity level (for the entire session as well as for subsessions) and to habituation (the expected decrease in activity with continuing time in the chamber). These parameters can change independently, for example two groups can show the same total activity levels, but very different patterns over time.

Results from the control animals should be examined carefully, looking at levels of variability, whether the average values are consistent with expectations at each age. If the control data appear reasonable, then the data from the treated animals should be examined for evidence of treatment-related effects. The following components of the test must be evaluated in concurrent control animals and compared to the historical control data from the same testing laboratory:

- Overall activity;
- Habituation; and
- Ontogeny

When analyzing the results from treated groups, it would be prudent to carefully examine the means and standard deviations presented in the report. It also would be clearer to graph results if the results indicated small changes relative to concurrent controls. The individual sessions should be examined carefully to detect any consistent changes from the controls. From the block data, habituation could be examined and percent change from the control should be indicated in the tabulated data. Habituation is often achieved by the slow and progressive decrease in motor activity as the number of block increases. Habituation to the testing procedures has also been shown indicating the necessity of testing the same animals at each time point including concurrent controls. It should also be noted that patterns observed within a session may indicate habituation while patterns observed between sessions may indicate evolution of activity over time Large variability in the data can mask effects and may bring into question the ability of the testing facility to detect treatment-related changes (Moser, 2000; Cory-Slechta et al. 2001; Raffaele et al, 2008; Tyl et al., 2008).

MA dataset should be presented as follows: **Total** (mean) values (Table 10) **and Sub-session** (Table 11) for both sexes at appropriate intervals. The mean with standard deviation should be presented along with the percent (%) change of controls.

Table 10. Mean (\pmSD) Motor Activity Data (Total Number of Movements/50 minutes)				
Test day	Dietary concentration (ppm)¹			
	0	LDT	MDT	LDT
Males				
PND 13				
PND 17				
PND 21				
PND 61				
Females				
PND 13				
PND 17				
PND 21				
PND 61				

Table 11. Mean (\pm SD) Sub-session Motor Activity in Males (# Movements/5 minute Sub-Session)				
Interval (min)	Dietary concentration (ppm)¹			
	Control	LDT	MDT	HDT
PND 13				
1-5				
6-10				
11-15				
16-20				
21-25				
26-30				
31-35				
36-40				
41-45				
46-50				
PND 17				
1-5				
6-10				
11-15				
16-20				
21-25				
26-30				
31-35				
36-40				
41-45				
46-50				
PND 21				
1-5				
6-10				

**Table 11. Mean (\pm SD) Sub-session Motor Activity in Males
(# Movements/5 minute Sub-Session)**

Interval (min)	Dietary concentration (ppm) ¹			
	Control	LDT	MDT	HDT
11-15				
16-20				
21-25				
26-30				
31-35				
36-40				
41-45				
46-50				
PND 60				
1-5				
6-10				
11-15				
16-20				
21-25				
26-30				
31-35				
36-40				
41-45				
46-50				

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

¹ = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

c. Auditory Startle

The auditory startle (AS) response is measured around the time of weaning (PND 21) and PND 60. The AS response is an apical test of sensorimotor function and is measured using either a force transducer or accelerometer. These devices provide different measure; a change in force on a platform (force transducer) or a movement of platform (accelerometer). The units of measure are quite different (grams vs. volts or arbitrary units). The auditory stimulus test is repeated at regular intervals for 50 trials (i.e, 5 blocks of 10 trials per session on each day of testing); normal response is decrease in amplitude over trials (habituation).

The equipment used measures the motor output of the acoustic-evoked motor reflex which begins approximately 8-10 ms after onset of the stimulus. Habituation of the response could be measured by using the same acoustic stimuli in multiple trials. The response could be proportional to the stimulus strength, frequency, and rise-time. It is also essential that experimental factors that could impact the quality of the data should be controlled. These factors could include the standardized stimulus parameters such as intensity, duration, frequency, as well as well-controlled environmental conditions, particularly noise levels. (Sheets et al, 1988; Sette et al., 2004; Tyl et al., 2008; Raffaele et al., 2008).

Auditory startle, like motor activity habituation, should be evaluated by examining the differences in response amplitude among treatment groups, both with respect to mean values for each block of trials and differences in habituation rate, as these endpoints may respond independently to nervous system toxicity.

Evaluation of the AS dataset must include the following: differences in mean values (total session and individual blocks) regardless of the presence statistical significance; differences in habituation, regardless of statistical significance of individual block differences; indication of data variability (e.g., based on variability, ontogeny, and gender-based comparisons); and review of the latency or time to peak responses. Confounding factors such as large variance, differences in baseline among groups, lack of habituation in controls (depending on age) and poor reporting (e.g., absence of overall mean response, absence of weight-adjusted individual data) should also be noted.

If the study results indicate a treatment-related effect on AS, then, it may be necessary to consider the results from other behavioral test (e.g., motor activity) to determine whether the results represent a specific auditory effect or an effect on motor capability. If additional information is not available to determine the specific nature of the change, then the findings should be interpreted as treatment-related effect on the function of the nervous system (Sette et al., 2004; Tyl et al., 2008; Raffaele et al., 2008)).

Results should be reported as mean response for 5 blocks of 10 trials each. For both males and females AS data should be presented as peak amplitude and latency to peak for all post natal days. The overall amplitude and latency data are presented in Table 12a. Interval amplitude and latency data are included in Tables 13. The mean with standard deviation should be presented along with the percent (%) change of controls.

Table 12. Mean (\pm SD) Overall (Blocks 1-5) Acoustic Startle Peak Amplitude (mv), Latency to Peak (msec) and Average Response Amplitude (mv)					
Dietary concentration (ppm)	Parameter	Males		Females	
		PND 20	PND 60	PND 20	PND 60
0	V _{MAX}				
	T _{MAX}				
	V _{AVE}				
LDT	V _{MAX}				
	T _{MAX}				
	V _{AVE}				
MDT	V _{MAX}				
	T _{MAX}				
	V _{AVE}				
HDT	V _{MAX}				
	T _{MAX}				
	V _{AVE}				

Table 13. Mean (\pmSD) Interval Acoustic Startle Peak Amplitude (mv), Latency to Peak (msec) and Average Response Amplitude (mv) In Males						
Dietary concentration¹ (ppm)	Parameter	1-10	11-20	21-30	31-40	41-50
PND 20						
0	V _{MAX}					
	T _{MAX}					
	V _{AVE}					
LDT	V _{MAX}					
	T _{MAX}					
	V _{AVE}					
MDT	V _{MAX}					
	T _{MAX}					
	V _{AVE}					
HDT	V _{MAX}					
	T _{MAX}					
	V _{AVE}					
PND 60						
MDT	V _{MAX}					
	T _{MAX}					
	V _{AVE}					
LDT	V _{MAX}					
	T _{MAX}					
	V _{AVE}					
MDT	V _{MAX}					
	T _{MAX}					
	V _{AVE}					
HDT	V _{MAX}					
	T _{MAX}					
	V _{AVE}					

V_{MAX} = maximum response amplitudeT_{MAX} = latency to V_{MAX}V_{AVE} = average response amplitude.

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

1 = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

d. Learning and Memory

Learning is defined as a relatively lasting change in behavior due to experience; and **Memory** is defined as the persistence of a learned behavior. Learning and memory (L&M) are evaluated around the time of weaning (PND 21) and PND 60. The test guideline does not specify the type of test to be used. However, based on the studies submitted to date, the most common tests used by testing laboratories are passive avoidance and various water mazes. These include: M-mazes and Y-mazes (shaped like their respective letters, rats placed in center corridors and have to turn right or left), Cincinnati maze or Biel maze (series of right-left turns) and Morris maze (round tank).

It should be noted that in the Morris maze, the rats have a free swim and have to find a submerged/hidden platform. All other mazes are submerged runways with an escape ramp at the end of one corridor and forced directional choices along the way. With the M- and Y-mazes, there is a single choice point (right or left). With the Biel and Cincinnati mazes, there are several successive choice points

Testing consists of two phases: Learning is the acquisition of the correct response; and memory is retention (memory) of the learned (correct) response. Depending on the method of testing used, data may be presented as time to completion of a task, number of trials to criterion, number of errors, or other measures. Testing on PND 60 should not be confounded by testing on PND 21. Therefore, it is imperative that different animals be used on these two time points or different tests should be performed at these two time points. The level of performance of the test used should be maintained such that both increases and decreases can be measured. Critical information necessary for proper evaluation of the study results include: the nature of the test system including its design and dimensions; understanding of the principal of the test procedures including the criteria to establish acquisition (learning) and retentions (memory); and experimental factors that can impact the quality of the data (Spear and Campbell, 1979; Stanton, 1994).

Learning is usually measured as a decrease in errors or latency to reach a goal, across trials or sessions. Retention is usually assessed by comparing the first trial during the retention session with the last trial of the learning session. The degree of retention will depend on the strength of the initial learning as well as the length of the delay between the final learning trial and the first retention trial; a decline in performance across these two trials would be attributable to a lack of retention. In all water maze tests, latency can be confounded by motor impairments, therefore, a test of a swim speed is very useful in interpreting latency data. In the absence of speed information, an effect on learning and memory is presumed.

Evaluation of datasets with runway-type mazes must include the following:

- Look at errors (wrong choices), latencies (which reflect both speed and learning), and some acquisition criteria (e.g., a specific number of consecutive correct response). Error criteria should be clearly defined as part of the study, report (i.e., do all 4 limbs have to be in the wrong selection area to be considered an error or would 2 limbs suffice).
- The number of errors in each trial should decrease with continued testing, latencies may decrease but asymptote (rats tend to swim at the same speed regardless of configuration), and animals should reach criteria after a similar number of trials (however, this can be variable).

NOTE: Average number of errors over all trial is not as useful as number of errors in each trial and does not allow appropriate evaluation of the rate of learning (acquisition curve).

- Memory (retention) can be measured by waiting some amount of time and repeating the test. ***NOTE: Last trial of training day should be compared to the first trial of retention day, therefore need data on last trial of acquisition phase.***
- Multiple trials on retention day (i.e., after the first trial that day) measures continued learning or re-learning.

Evaluation of datasets with the Morris water maze should include the following:

- Look at latencies (which reflect learning the position of the platform), and other measures are useful (e.g., direction of swimming at the start of the trial, spatial distribution of the search strategy swim path)
- Latencies should decrease greatly over training, with low variability at end of training
- Probe trials are conducted for memory tests—the platform is removed and time spent in the area where it should be is measured.

Learning and memory data with mean and standard deviation should be presented for all session as shown in Tables 14, 15 and 16. Data for all trials (individually or in blocks) should be presented.

Table 14. Passive Avoidance Performance (mean \pm SD).					
Session/Parameter		Dietary concentration (ppm) ¹			
		Control	LDT	MDT	HDT
Males					
Session 1 (Learning)	Trials to criterion				
	Latency Trial 1 (sec)				
	Latency Trial 2 (sec)				
	Failed to learn				
Session 2 (Memory)	Trials to criterion				
	Latency Trial 1 (sec)				
	Latency Trial 2 (sec)				
Females					
Session 1 (Learning)	Trials to criterion				
	Latency Trial 1 (sec)				
	Latency Trial 2 (sec)				
	Failed to learn				
Session 2 (Memory)	Trials to criterion				
	Latency Trial 1 (sec)				
	Latency Trial 2 (sec)				

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

¹ = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

Table 15. Water Maze Percentage of Successful Trials

Cut-off	Control	LDT	MDT	HDT	Control	LDT	MDT	HDT
	Males				Females			
Day 21 - Learning phase								
3 sec								
4 sec								
5 sec								
6 sec								
7 sec								
8 sec								
9 sec								
10 sec								
Day 24 - Memory phase								
3 sec								
4 sec								
5 sec								
6 sec								
7 sec								
8 sec								
9 sec								
10 sec								
Day 59 - Learning phase								
3 sec								
4 sec								
5 sec								
6 sec								
7 sec								
8 sec								
9 sec								
10 sec								
Day 62 - Memory phase								
3 sec								
4 sec								
5 sec								
6 sec								
7 sec								
8 sec								
9 sec								
10 sec								

Table 16. Mean Percentage of Successful Trials at 1.5× Straight Channel Swim Time

Interval (phase)	Control	LDT	MDT	HDT
Males				
Day 21 (learning)				
Day 24 (memory)				
Day 59 (learning)				
Day 62 (memory)				
Females				
Day 21 (learning)				
Day 24 (memory)				
Day 59 (learning)				
Day 62 (memory)				

4. Postmortem

The postmortem evaluation includes a macroscopic evaluation of the brain and tissues of the nervous system, brain weight parameters, and gross morphometry of the brain, neurohistological examinations and quantitative analyses of the major areas of the brain.

A brief overview for neuropathology evaluation is presented here. For a comprehensive guidance on the evaluation of developmental neuropathology, the reader is advised to consult the following journal articles:

- Bolon, B, Garman, R, Jenses, K et al (2006). A “Best Practices” approach to neuropathologic assessment in developmental neurotoxicity testing-for today. *Toxicologic Pathology* 34: 296-313
- Duffell, SJ, Soames, AR and Gunby, S (2000). Morphometric analysis of the developing rat brain. *Toxicologic Pathology*, 28 (1) 157-163
- Garman, R, Fix A, and Jortner et al (2001). Methods to identify and characterize developmental neurotoxicity for human health risk assessment: II. Neuropathology. *Environmental Health Perspectives* 109. 93- 100.
- Jensen, K.F.,and S. Catalano (1998) Brain morphogenesis and developmental neurotoxicology. In *Handbook of Developmental Neurotoxicology*. Chapter 1. W. Slikker and L.W. Chang, Eds. Academic Press. Pp. 3-41.
- Kaufman W and Groters S (2006). Developmental neuropathology in DNT studies-A sensitive tool for the detection and characterization of developmental neurotoxicants . *Reproductive Toxicology* 22. 196-213.

a. Brain weights

Postfixation brain weights are obtained on both juveniles (PND 11) and adults (PND 70). Changes in brain weight in developing animals are cause for concern because brain weight is generally protected during malnutrition or weight loss unlike many other organs or issues. This phenomenon is commonly referred to as “brain sparing”. It is not appropriate to express brain weight changes as brain-to-body weight ratios (*i.e.*, relative weights) and thereby dismiss the changes in absolute brain weights.

Statistically significant decreases in absolute brain weights seen in the pups in the absence of any other changes in the pups are a concern and regarded as attributable to treatment. Similarly, alterations in brain weight in the presence of decreased in pup body weight should also be considered as adverse effects and it is not necessary to determine its relations to body weight changes which presumably are also adverse. Although alteration in brain weight may not be a sensitive indicator of impaired neurological development, when present it can be an indicator of impaired brain development in the context of DNT testing.

In order to ascertain the impact of the findings, the reviewer should consider the following factors: statistical significance identified in groups comparisons; biological significance; percent change from control; dose-related trends, background (historical) incidences in the testing laboratory; and consistencies or inconsistencies in patterns of response. In general, the concurrent control data should be the basis for examining the data of treated groups for compound-related effects. In some incidences, if the concurrent control data raises questions or concerns, then the recent historical control data from the testing laboratory (in combination with the concurrent control data) could be used in aiding interpretation of unusual or inconsistent findings in brain weight data. As shown in Table 17, the mean with standard deviation should be presented along with the percent (%) change of controls.

Table 17. Brain (\pmSD) Brain Weights				
Parameter	Dietary concentration (ppm) ¹			
	0	LDT	MDT	LDT
Males				
Day 21				
Terminal body weight (g)				
Brain weight (g)				
Day 72				
Terminal body weight (g)				
Brain weight (g)				
Females				
Day 21				
Terminal body weight (g)				
Brain weight (g)				
Day 72				
Terminal body weight (g)				
Brain weight (g)				

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

¹ = If the test compound is administered by gavage, the unit should be presented in mg/kg/day

b. Neuropathology

Qualitative neuropathological examinations of samples from all major **brain regions** of offspring at **PND 21** and termination (**PND 60**) and **peripheral tissues in adults only** are conducted to:

- 1) Identify **regions** within the nervous system exhibiting evidence of neuropathological changes;
- 2) Identify **types** of neuropathological lesions; and
- 3) Determine the range of **severity** of the observed neuropathological alterations.

The Guideline requirement includes histopathological examination of tissues of the developing nervous system at two different time points. The treatment time points are extremely important due to the special vulnerability of the nervous system during defined windows of development.

Specifically, the guideline requires morphologic assessments of the peripheral nervous system (peripheral nerves and ganglia, spinal nerve roots and possibly other structures) of the adult rats. Tissues from control and high dose groups should be examined. If treatment-related findings are seen at the high dose, then tissues of animals from the low and mid-dose tissues should be examined. Also, if alterations are seen, “blind” evaluation is required. All neuropathological alterations should be assigned a subjective grade indicating severity. A hematoxylin and eosin stain may be sufficient for evaluation, but subject to the pathologist’s judgment and the kind of alterations observed, other stains may be considered appropriate to identify and characterize particular types of alterations.

Various types of histological changes can occur in the brain tissue of offspring following maternal exposure to chemicals with neurotoxic potential. For example, cellular alterations can manifest as neuronal vacuolation, degeneration and necrosis (cell death) while tissues changes can be characterized as astrocyte proliferation, leukocyte infiltration and cystic formation. Structural changes indicative of developmental insult can occur as gross changes in size or shape of the brain regions, death of neuronal precursors, abnormal proliferation, or migration; abnormal differentiation and evidence of hydrocephalus. Specific changes in nerve cell bodies may include chromatolysis, vacuolation, and necrosis. Axons can undergo swelling, degeneration, and atrophy while myelin sheath changes can include folding, edematous splitting and demyelination. (Garman et al., 2001; Kaufman and Groeters, 2006; Bolon, et al., 2006).

In DNT-studies, neuropathological examinations include an examination of the damage to the developing fetal nervous system. Consequently, it is essential that treatment-related changes be distinguished from normal developmental events known to occur at a certain stage of the development corresponding to the time of sacrifice. Examples of notable changes indicative of developmental insult include, but are not limited to (OECD, 2007):

- Alterations in the gross size or shape of the olfactory bulbs, cerebellum and cerebrum;
- Alterations in the relative size of various brain regions, including decreases or increases in the size of regions resulting from the loss or persistence of normally transient populations of cells or axonal projections (e.g., external germinal layer of cerebellum, corpus callosum);
- Alterations in proliferation, migration, and differentiation, as indicated by areas of excessive apoptosis or necrosis, clusters or dispersed populations of ectopic, disoriented or malformed neurons or alterations in the relative size of various layers of cortical structures;
- Alterations in patterns of myelination, including an overall size reduction or altered staining of myelinated structures;
- Evidence of hydrocephalus, in particular enlargement of the ventricles, stenosis of the cerebral aqueduct and thinning of the cerebral hemisphere

c. Morphometrics

Quantitative brain measurements (morphometric) provide data on the detection of the disruption of the developmental process that can be observed in rate or extent of growth particular to the brain regions. Morphometric data can also be valuable in the interpretation of treatment-related differences in brain weight or brain morphology (DeGroot et al., 2005a, 2005b). A minimum of three representative locations should be examined: neocortex, hippocampus and cerebellum. Additional areas (e.g., basal ganglia, thalamus, hypothalamus) may need to be examined based on chemical-specific data or information. Basically there are three types of morphometric data: linear measurement, areal measurements, and cell profile counts. Brain sections used for linear measurements must be highly homologous among animals on a given study to have predictive value. Linear or areal measurements require the use of homologous sections carefully selected based on reliable microscopic landmarks. The guideline requires an estimation of the thickness of the major layers at representative locations within the neocortex, hippocampus, and cerebellum. Linear measurements should be taken bilaterally and recorded separately, even if

mean values for the right-and left-sided measurements are used for statistical analysis. (Duffel et al., 2000; Bolon et al, 2006; USEPA, 1998a, OECD, 2007).

Evaluation of the data submitted to the Agency shows that morphometric data of all three regions provide important information about brain pathology, and that no one measure is reliably predictive of effects on the others (Raffaele, 2005). In addition, Agency experience suggest that it may be preferable to evaluate all treatment groups at the same time, to allow for interpretation of neuropathology data in the context of the contemporaneous analysis of all the data from the other groups. It also avoids delays in study reviews that could result if questions arise about the need for lower dose evaluations. At a minimum, tissues from all dose groups should be taken to the block stage.

Morphometric data with mean and standard deviation should be presented along with the percent (%) change of controls as shown in Table 18.

Table 18. Mean (±SD) Morphometric data								
Parameter	Dietary concentration (ppm)							
	0	LDT	MDT	HDT	0	LDT	MDT	HDT
	Males				Females			
PND 12								
Cerebrum (mm)								
Cerebellum (mm)								
Frontal Cortex (·m)								
Parietal Cortex (·m)								
Corpus Striatum (·m)								
Corpus Callosum (·m)								
Hippocampal Gyrus (·m)								
Cerebellum Height (·m)								
Ext. Germinal Layer (·m)								
Termination								
Cerebrum (mm)								
Cerebellum (mm)								
Frontal Cortex (·m)								
Parietal Cortex (·m)								
Corpus Striatum (·m)								
Corpus Callosum (·m)								
Hippocampal Gyrus (·m)								
Cerebellum Height (·m)								

LDT = Low Dose, MDT = Mid Dose, HDT = High Dose

C. Statistical Analyses

All results should be analyzed using statistical models appropriate to the experimental design. For the selection of appropriate statistical methods and data analyses, consultation with statisticians familiar with the appropriate analyses of the data is recommended. The evaluation should include the relationship between the doses of the test substance and the presence or absence, incidence, and extent of any neurotoxic effect. The evaluation should include appropriate

statistical analyses with the basic premise that the litter is the primary unit of analysis. The choice of a parametric or non-parametric analysis should be justified by considering factors such as the nature of the data (transformed or not) and their distribution, as well as relative robustness of the statistical analysis selected. Statistical analysis should be used as tool that guides the data rather than the means to interpret the data. Statistical significance does not necessarily signify biological significance and scientific judgment and relevant historical control data should be used to distinguish between fortuitous and real findings (USEPA 1998a, OECD, 2007; Holson et al., 2007; Tyl et al., 2008).

Holson et al., 2007 reviewed many key considerations in the analyses of the DNT studies and made the following recommendations:.

- Evaluate the data and not just the level of significance. Wherever possible the graphing of data is encouraged;
- Include sex and sex-by-treatment interaction as factors in the analysis of all dependent variables collected on both sexes;
- Litter must remain a factor in analysis throughout the study, not just in young animals. If each litter contributes animals of both sexes, then sex and the sex by treatment interaction must be analyzed as a correlated variable;
- To assess adaptation or changes that occur over time, repeated-measures methodologies should be utilized to evaluate the effects of different dose groups while accounting for the correlated data resulting from multiple measurements on the same animal;
- Clearly indicate the data to be provided by each procedure. Do not test hypotheses on the same data that generated them. Always test them in a new data set;

- Identify the type of data that each endpoint represents (e.g., body weight is a continuous endpoint, degree of lacrimation is an ordinal endpoint) and utilize the most appropriate statistical methodology for that type. Describe in detail all the statistical analyses in the protocol and the study report;
- Running a statistical analysis after seeing the data can only generate hypotheses to be confirmed, but no conclusions;
- Consider strategies to address the multiplicity problem in the study, or at least indicate how the multiplicity problem will be addressed in the study. Use complex multifactorial statistical analyses to substantially reduce multiplicity of significance tests. This approach has the added benefit of allowing tests of interaction terms which are not addressed in simpler designs;
- Provide the total count of derived p-values (significant and nonsignificant). Preferably, report exact p-values with their associated F values and degrees of freedom, as appropriate. If p-values are adjusted for multiple comparisons, unadjusted p values should also be provided and the method of adjustment should be identified;
- Address the strength of association between dependent and independent variables;
- Use pairwise comparison procedures that are optimal for the questions being addressed (e.g., use Dunnett's if the only concern is to compare treatment mean values to control); and
- Consider the use of statistical methodology specifically designed for censored data when the data include a substantial number of such measurements (e.g., passive avoidance latencies in which crossover never occurs).

D. Use of Control Data

1. Concurrent (vehicle) Control

Concurrent control data are required for every study. Comparison of concurrent study control data with the data from treated animals should always take precedence over comparison with historical control data. Evaluation of the differences between concurrent (vehicle) control and treated groups should include a review of the nature of the control data. Some data (e.g., quantile data with a normal value near zero) may lead to small changes to be statistically different from control when they are in reality indistinguishable one from another. Similarly, precision of equipments used or endpoint measuring techniques (e.g., field sensor to measure motor activity) must also be evaluated since imprecision may result from human and/or instrument factors such as reproducibility, technology and bias.

2. Historical Control

Historical control data, which are generally comprised of well-characterized negative (vehicle) control data from multiple studies, can be a useful tool for interpreting study findings. Historical control data can sometimes provide a guide for determining the biological significance of a statistically significant difference observed in the study.

Historical control data may be used to identify aberrant control values, in order to determine if the results of the concurrent control group is consistent with the larger population of controls or if it is atypical. If a determination is made that the value for a concurrent control is atypical, individual animal data should be examined to determine if high or low values fall outside of the historical control range and are a source of bias in calculating the mean. However, aberrant concurrent control data raise serious concerns regarding the validity of the study, and a determination that concurrent controls are aberrant may indicate that the study may be invalid.

Historical control data can also be used to understand the relevance of a low- or high-incidence finding. For low incidence data, knowledge of the range of historical control value is important to differentiate a genuine effect from a false difference from the concurrent control. By the nature of the low incidence, it is possible for a treated group to show a low spontaneous incidence, yet the control group has lower or even no incidence. In these cases, historical control data will provide data on the overall spontaneous occurrence of the finding. If the treated group incidence falls within the larger population range, the difference from the concurrent control may in some incidences be considered not to be an effect. For high incidence findings, historical control data will provide data on what range is considered normal for the species under the test.

Historical control data should only include studies conducted within an appropriate time period on either side of the study under review, conducted in the same strain, age, sex of experimental animals obtained from the same supplier, should be from the same conducting laboratory and should be reasonably of the same period to the study under evaluations. Also, it is important to verify that the test conditions and procedures were identical to those in the study under evaluation. In cases where procedures varied or information is unavailable, historical control data should not be used. Overall, the interpretation and use of historical control data requires careful considerations, and the application of scientific judgment and expertise.

3. Positive Control

Periodic verification is needed to demonstrate that the test methods and the manner in which they were conducted can adequately and reliably detect the target effect(s) across a range of exposures to known positive control agents. For such quality control, known neurotoxic agents (positive controls) and toxic agents with no neurotoxic potential (negative controls) should be generated in specifically designed studies prior to conduct of the DNT. Positive control studies may require different chemicals for different effects (behavioral, physiological, neuropathology) and at several doses. This verification is of particular importance in aiding interpretation of the test data. Positive control data do not have to be from studies using prenatal exposures. However, the laboratory must demonstrate competence in evaluation of effects in neonatal animals perinatally exposed to chemicals and establish test norms for the appropriate age group.

All behavioral positive control data should be able to detect increases or decreases in measured parameters. Motor activity positive control data should demonstrate the ability to detect both increases and decreases in motor activity. Auditory startle data should demonstrate the ability to detect an increase or decrease in the response with a possible change in habituation and a lack of response to a stimulus. Learning and memory tests should demonstrate a change in the ability to learn or a change in retention of the learned task. Pathology positive control data should demonstrate the ability to detect central and peripheral nervous system pathology (separate groups may be used to demonstrate each type of pathology, e.g., acrylamide for peripheral nervous system pathology and trimethyl tin for central nervous system pathology including changes in nervous system development).

The methods for each endpoint should be completely described, and must be the same as those used in the study being evaluated (e.g., the same equipment should be used, motor activity sessions should be of the same duration, the observation arena should be the same, the same number of animals should be used, the same sections should be evaluated for neuropathology, using the same types of stains, etc.), and preferably the same personnel should have conducted the testing. If different personnel are conducting the testing, data should also demonstrate inter-observer reliability for the FOB (i.e., the same results should be seen regardless of who is doing the observations).

The data presentation should be complete enough to evaluate the sensitivity of the method, including individual data and measures of variability. Statistical evaluations used to demonstrate sensitivity should also be the same as those used in the study being evaluated. The number of animals per test group should not be greater than that used in the study under evaluation. The positive control data should have been collected within a reasonable time frame before the current study (e.g., the last few years). New data should also be collected when observational personnel, procedures, or other critical laboratory elements change.

To date, considerable variability has been found in the quality of positive control data submitted from laboratories conducting DNT studies. Common problems with positive control data include faulty study design, inadequate reporting of the results, and problems with the manner in which the data were reported (Raffaele et al 2002; Crofton et al. 2004).

V. DEVELOPMENTAL NEUROTOXICITY HAZARD CHARACTERIZATION

The DNT is designed to assess the adverse effects of *in utero* and postnatal exposure on the development of and function of the nervous system and to provide dose-response characterization of those outcomes. Many of the results in a DNT study are an evaluation of multiple measures of different endpoints over doses and time; these aspects of the study design allow measurement of latent, transient, progressive and persistent effects. Critical periods exist both during prenatal and postnatal periods in which disruption of functional competence can occur and the effect of a toxicant is likely to vary depending on the time and degree of exposure (Rodier, 1978, 1979, 1986, 1990, 1994, 1995). Therefore, effects are important at any time point they are observed and the effect may be observed in the pattern of response. In addition, differences in response by sex can occur. The focus of the DNT study, therefore, is detailed/extensive evaluation of the offspring with limited evaluation of the dams.

Evaluation of this study involves examining all available data and the associated dose, routes, timing, and durations of exposure to determine qualitatively if a test chemical causes developmental neurotoxicity. Scientific judgment and expertise will be required in the interpretation of the DNT data due to the complex interrelationship among study design, experimental procedures, the number and types of endpoints affected, adequacy of the doses tested, and statistical analyses etc. The occurrence, detection, and interpretation of adverse effects may be influenced by the mode of administration (i.e., gavage vs. dietary).

A series of reports have also emphasized the importance of using well-designed studies and well-trained personnel in order to achieve reliable results. The methodologies and results from DNT studies submitted to the Agency have found large differences and variability of control data from different testing laboratories, for a number of measured endpoints (Crofton et al., 2001; Crofton et al. 2004; Raffale et al., 2003, 2004, 2005, 2006, 2008; Makris et al. 2005, 2009). These findings stress the importance of careful evaluation and interpretation of the results from each study with respect to the appropriateness of the procedures used their sensitivity in detecting the effects of concern, and the quality of the resulting data.

Considerations of the toxicity database, (especially the prenatal developmental and the reproduction studies in rats) in its entirety can provide information that could place the DNT findings in perspective. Consequently, a weight-of-evidence approach using data from all available studies should be used in the interpretation of data, hazard characterization, dose response assessments, and hazard identification for human health risk assessments.

The overall hazard characterization of effects from the DNT study should be based on the 1) integrated assessment of functional (neurological, behavioral, and physiological) and neuropathological changes, using appropriate biomarkers of effect and 2) on the well characterized exposures (dose, duration, route, biomarkers of exposures etc).

An important aspect of hazard characterization in the DNT study is that effects in the offspring should be considered to be of concerns and not be negated by maternal toxicity or by alterations in pup growth or development. This is the same approach as that taken for the pre-natal developmental toxicity studies in which developmental effects are considered to represent developmental toxicity and are not discounted as secondary to maternal toxicity. The approach that any indication of toxicity to the offspring should be considered as an offspring-specific, treatment related effect, provides a conservative interpretation for public health protection.

Dose-response analysis is a critical part of the qualitative characterization of a chemical's potential to produce developmental neurotoxicity and involves the description of the dose-response relationship in the available data. Dose-response assessment includes determination of the no-observed-adverse-effect level (NOAEL) and the lowest-observed-adverse-effect level (LOAEL) for the most sensitive endpoint, whether it is in the dams (maternal) or offspring, and whether or not it is specifically neurotoxic or developmentally neurotoxic. In addition to establishing a Point of Departure (POD) such as a NOAEL, LOAEL, or calculating a BMD, the dose-response evaluation defines the range of doses that are developmentally neurotoxic for a given chemical, sex, duration, and endpoint. For example, for behavioral assessments, it is important to keep in mind that the dose response curve may exhibit not only monotonic but also non-linear (including U- or inverted U-shaped) functions. In addition to considering the shape of the dose response curve, it should also be recognized that neurotoxic effects can vary in terms of nature and severity across dose and duration.

In general, evaluation of the findings in the DNT involves interpretive issues that are common across all endpoints and can include the following (Crofton et al, 2001; Milesen et al., 2001; Cory-Slechta et al., 2001; Garman et al., 2001; Dorman et al., 2001; OECD, 2007; Tyl et al, 2008)

- Absence/presence/shape of the dose-response curve;
- Consistency over time (the temporal characteristics of a treatment-related effect are essential for determining the biological relevance of the observed effects);
- Consistency across doses (the shape of the dose-response curves such as linear, threshold, flat ("hockey stick") or non-linear (including U-or inverted U-shaped);
- Consistency (or not) between genders (changes in one parameter that are evident in one gender only should be carefully evaluated for biological significance);
- Consistency across parameters (e.g., motor activity along with learning and memory);
- Atypical control baselines (e.g., lack of auditory startle habituation in motor activity or startle task, or inadequate associative learning and memory);
- High variability in potential treatment-related effects;
- Use of positive controls to test and confirm (a positive control chemical should disrupt the same neurobiological system as the test chemical, if it is known *a priori*);
- Evaluation of data that includes a discussion of both biological and/or statistical significance;
- Evaluation of data that includes a discussion on the potential differences in toxicity due to gavage vs. dietary mode of administration
- The animal model is sensitive to the DNT effects of interest ; and
- The laboratory technicians are trained and competent to detect the various endpoints measured in the DNT.

As in all hazard characterization, the available toxicity database should be considered in its entirety. Considering the results of all of the studies together provides a context which can enhance understanding of the overall toxicity profile and hazard identification for that chemical. Key in this process is consideration of the strengths and weaknesses in the database as well as identification of the nature of any uncertainties in the data set. When a DNT study is available, it is included in this overall weight-of-evidence consideration of hazard identification and dose/endpoint selection for human health risk assessment.

To date, the Agency has received 75 DNT studies and reviews have been completed for 58 of these studies. Out of the 58 completed reviews, the DNT study was selected for 8 chemicals as the critical study for dose/endpoint selection for risk assessment; 4 were used for establishing the acute Reference Dose (aRfD) and 4 were used for establishing the chronic RfD. For four of these eight DNT studies, the critical effects either included or were solely based upon offspring behavioral and neuropathologic parameter. A single study was sometimes used for multiple risk assessment scenarios. An additional 17 cases were identified where an endpoint of concern from a DNT study could potentially be selected for use in one or more scenarios in future risk assessment actions (Rowland et al, 2007).

The Data Evaluations Records (DERs) for submitted DNT studies and final HED risk assessment documents that describe the consideration of the DNT study in endpoint selection are available in the Developmental Neurotoxicity Discussion and the OPPIN Risk assessment Documents Databases (RADD) in LotusNotes, respectively. In addition, critical journal articles, guidance documents and SEPs referenced herein are also available in the DNT Discussion database.

VI. DATA EVALUATION REPORT

Once the study has been evaluated for acceptability and interpreted using the principles described in previous sections of this SEP, a Data Evaluation Record (DER) must be prepared. Guidance for the DER preparation is available on HED Policy Documents Database in Lotus Notes.

VII. REFERENCES

- Bolon, B, Garman, R, Jenses, K et al (2006). A "Best Practices" approach to neuropathologic assessment in developmental neurotoxicity testing for today. *Toxicoc Pathology* 34: 296-313
- Campbell, B.A. and Haroutunian, V. (1981) Effects of age on long-term memory: Tetention of fixed interval responding. *J. Gerontol.* 36:338-341.
- Chapin, RE., Harris, M.W., Davis, B.J., Ward, S.M., Wilson, R.E., Mauney, M.A., Lockhart, A.C., Smialowicz, R.J., Moser, V.C., Burka, L.T., and Collins, B.J. (1997) The effects of perinatal/juvenile methoxychlor exposure on adult rat nervous, immune, and reproductive system function. *Fundam. Appl. Toxicol.* 40:138-157.
- Cory-Slechta, D.A., Crofton, K.M., Foran, J.A., Ross, J.F., Sheets, L.P., Weiss, B., and Milesen, B. (2001) Methods to identify and characterize developmental neurotoxicity for human health risk assessment. I: Behavioral effects. *Environ. Health Perspect.* 109:79-91.

- Crofton, K.M., Makris, J.L., Sutton, S.L., Makris, Raffaele, K.C., Sette, W.F (2001). Developmental neurotoxicity testing guidelines: variability in morphometric assessments of Neurotoxicology. Toxicologist 2001. Crofton, K.M., Makris, S.L., Sette, W.F., Mendez, E., and Raffaele, K.C. (2004) A qualitative retrospective analysis of positive control data in developmental neurotoxicity studies. Neurotoxicol. Teratol. 26:345-352.
- De Groot, D. M.G, Bos-Kuijers, M.H.M, Kaufman, W.S.H., Lammeres, J.H.C.M et al., 2005a. Regulatory developmental neurotoxicity testing: A model study focusing on conventional neuropathology endpoints and other perspectives. Environ.Toxicol. Pharmacol. 19:745-755.
- De Groot, D. M.G, Bos-Kuijers, M.H.M, Kaufman, W.S.H., Lammeres, J.H.C.M et al., 2005b. 2D and 3D assessment of neuropathology in rat brain after prenatal exposure to methylazoxymethanol, a model for developmental neurotoxicity. Reprod. Toxicol. 20:417-432.
- Dorman, D.C., Allen, S.L., Byczkowski, J.Z., Claudio, L., Fisher, Jr., J.E., Fisher, J.W., Harry, G.J., Li, A.A., Makris, S.L., Padilla, S., Sultatos, L.G., and Milesen, B.E. (2001) Methods to identify and characterize developmental neurotoxicity for human health risk assessment. III: Pharmacokinetic and pharmacodynamic considerations. Environ. Health Perspect. 109:101-111, Suppl. 1.
- Duffell, SJ, Soames, AR and Gunby, S (2000). Morphometric analysis of the developing rat brain. Toxicologic Pathology, 28 (1) 157-163
- Friede, R.L. (1975) Developmental Neuropathology. New York: Springer-Verlag. pp. 1-23, 397-313, 326-351.
- Garman, R, Fix A, and Jortner et al (2001). Methods to identify and characterize developmental neurotoxicity for human health risk assessment: II. Neuropathology. *Environmental Health Perspectives* 109. 93- 100.
- Gerrish, C.J., Onischak, C.M., and Alberts, J.R. (1998) Acute, early thermal experience alters weaning onset in rats. Physiol. Behav. 64:464-474.
- Green, R.J. and Stanton, M.E. (1989) Differential ontogeny of working memory and reference memory in the rat. Behav. Neurosci. 103:98-105.
- Hanley, Jr., T.R., and Watanabe, P.G. (1985) Measurement of solid feed consumption patterns in neonatal rats by ¹⁴¹Ce-radiolabeled microspheres. Toxicol. Appl. Pharmacol. 77:496-500.
- Harrington, M.E., and Coscina, D.V. (1983) Early weight gain and behavioral responsivity as predictors of dietary obesity in rats. Physiol. Behav. 30: 763-770
- Holson, R., Freshwater, L., Maurissen, J.P., Moser, V.C., and Phang, W. Statistical Issues and Techniques Appropriate for Developmental Neurotoxicity Testing. A Report from the ILSI Research Foundation/Risk Science Institute Expert Working Group on Neurodevelopmental Endpoints. Neurotoxicology and Teratology; Vol 30, Issue 4, July-August 2008, P326-348.

Ison, J.R. (1984) Reflex modification as an objective test for sensory processing following toxicant exposure. *Neurobehav. Toxicol. Teratol.* 6:437-445.

Jensen, K.F., and S. Catalano (1998) Brain morphogenesis and developmental neurotoxicology. In *Handbook of Developmental Neurotoxicology*. Chapter 1. Slikker and L.W. Chang, Eds. Kaufman W and Groters S (2006). Developmental neuropathology in DNT studies-A sensitive tool for the detection and characterization of developmental neurotoxicants . *Reproductive Toxicology* 22. 196-213.

Khera, K., Grice, H.C., and Clegg, D. (1989) Interpretation and extrapolation of reproductive data to establish human safety standards. *Current Issues in Toxicology*. New York: Springer-Verlag. p. 29.

Korenbrot, C.C., Huhtaniemi, I.T., and Weiner, R.I. (1977) Preputial separation as an external sign of pubertal development in the male rat. *Bio. Reprod.* 17:298-303.

Krasnegor, N.A., Blass, E.M., Hofer, M.A., and Smotherman, W.P. (Eds.) (1987) *Perinatal Development: A Psychobiological Perspective*. Orlando: Academic Press. pp. 11-37, 145-167.

Kucharski, D. and Spear, N.E. (1984) Conditioning of aversion to an odor paired with peripheral shock in the developing rat. *Develop. Psychobio.* 17:465-479.

Luna, L.G. (Ed.) (1968) *Manual of Histologic Staining Methods of the Armed Forces Institute of Pathology*, 3rd ed. New York: McGraw-Hill. pp. 32-46.

Makris, S.L, Raffaele, K., Sette, W., and Seed, J. (1998) A retrospective analysis of twelve developmental neurotoxicity studies submitted to the USEPA Office of Prevention, Pesticides, and Toxic Substances (OPPTS). Draft 11/12/98.

Makris, S., Raffaele, K., Allen, S., Bowers, W.J., Hass, U., Alleva, E., (2009) A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426. *Environ. Health. respect.* 117(1): 17-25

Miller, D.B. and Eckerman, D.A. (1986) Learning and memory measures. In: *Neurobehavioral Toxicology*, Z. Annau, Ed. Baltimore: Johns Hopkins University Press. pp. 94-149.

Mose, V.C. (2000) The functional observational battery in adult and developing rats. *Neurotoxicol.* 21:989-996.

Moser, V.C., Chanda, S.M., Mortensen, S.R., and Padilla, S. (1998) Age- and gender-related differences in sensitivity to chlorpyrifos in the rat reflect developmental profiles of esterase activities. *Toxicol. Sci.* 46:211-222.

Moser, V.C., Barone, Jr., S., Smialowicz, R.J., Harris, M.W., Davis, B.J., Overstreet, D., Mauney, M., and Chapin, R.E. (2001) The effects of perinatal tebuconazole exposure on adult neurological, immunological and reproductive function in rats. *Toxicol. Sci.* 62:339-352.

- OECD (2007) OECD Guideline for the Testing of Chemicals - 426 Developmental Neurotoxicity Study. Organisation of Economic Cooperation and Development, October, 2007.
- OECD (2008) OECD Series on Testing and Assessment, Nuber 43: Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment. Organisation for Economic Co-operation and Development. ENV/JM/MONO(2008)16, 24-July-2008, Paris.
- Raffaele, K.C., Makris, S.L., Sette, W.F., and Crofton, K.M. (2002) Developmental neurotoxicity testing guidelines: a qualitative retrospective analysis of positive control data. *Toxicologist* 66:134.
- Raffaele, K.C., Sette, W., Makris, S.L., Moser, V., and Crofton, K.M. (2003) Motor activity in developmental neurotoxicity testing: A cross-laboratory comparison of control data. *Toxicologist*, 72, 123.
- Raffaele, K.C., Gilbert, M.E., and Crofton, S (2004) Learning and memory tests in developmental neurotoxicity testing: A cross-laboratory comparison of control data. *Toxicologist*, 78, 276
- Raffaele, K.C., Sette, W., Doherty, J.D., Makris, S.L., and Crofton, K.M. (2005) Neuropathological findings in developmental neurotoxicity testing: comparison of qualitative and quantitative evaluations. *Toxicologist*, 85:200
- Raffaele, K.C., Fisher, J.E., Hancock, S., Hazelden, K., and Sobrian, S.K. (2008) Determining normal variability in a developmental neurotoxicity test. A report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. *Neurotoxicology and Teratology* 30: 288-325.
- Ralis, H.M., Beesley, R.A., and Ralis, Z.A. (1973) *Techniques in Neurohistology*. London: Butterworths. pp. 57-145.
- Rodier, P.M. (1978) Behavioral teratology. In: *Handbook of Teratology*, vol. 4, J.G. Wilson and F.C. Fraser, eds. New York: Plenum Press. pp. 397-428.
- Rodier, P.M. and Gramann, W.J. (1979) Morphologic effects of interference with cell proliferation in the early fetal period. *Neurobehav. Toxicol.* 1:129-135.
- Rodier, P.M. (1986). Time of exposure and time of testing in developmental Neurotoxicology. *NeuroToxicology* 7(2) 69-76.
- Rodier, P.M. (1990). Developmental neurotoxicology. *Toxicol. Pathol.* 18:89-95.
- Rodier, P.M (1994) Vulnerable periods and processes during central nervous system development. *Environ. Health Perspect.* 102 (Suppl 2): 121-124.
- Rodier, P.M. (1995) Developing brain as a target of toxicity. *Environ. Health Perspect.* 103 (Suppl 6): 73-76.

- Rowland, J, Makris, S, Raffaele, K, Shumacher, K, and Scarano, L. 2007. A retrospective review of the use of the developmental neurotoxicity study in pesticide assessments. *Toxicologists* 96:369
- Ruppert, P.H., Dean, K.F., and Reiter, L.W. (1985) Development of locomotor activity of rat pups in figure-eight mazes. *Develop. Psychobiol.* 18:247-260.
- Sette, W.F., Crofton, K.M., Makris, S.L., Doherty, J.D., and Raffaele, K.C. (2004) Auditory startle reflex habituation in developmental neurotoxicity testing: A cross-laboratory comparison of control data. *Toxicologist*, 78, 275.
- Sheets, L.P., Dean, K.F., and Reiter, L.W (1988) Ontogeny of the acoustic startle response and sensitization to background noise in the rat. *Behav. Neurosci.* 102: 706-713
- Spear, N.E. and Campbell, B.A. (Eds.) (1979) *Ontogeny of Learning and Memory*. New Jersey: Erlbaum. pp. 101-133, 157-224.
- Stanton, M.E. (1994) Assessment of learning and memory in developmental neurotoxicology. In: *Developmental Neurotoxicology*, G.J. Henry, Ed. Boca Raton: CRC Press. pp. 123-155.
- Suzuki, K. (1980) Special vulnerabilities of the developing nervous system to toxic substances. In: *Experimental and Clinical Neurotoxicology*, P.S. Spencer and H.H. Schaumburg, Eds. Baltimore: Williams and Wilkins. pp. 48-61.
- Tyl, R.W, Crofton, K, Moretto, A, Moser, V, Sheets, L, and Sobotka, T.J (2008) Identification and interpretation of developmental neurotoxicity effects. A report from the ILSI Research Foundation / Risk Science Institute expert working group on neurodevelopmental endpoints. *Neurotoxicology and Teratology*. 30: 348-381
- USEPA (Environmental Protection Agency) (1985) Hazard evaluation division standard evaluation procedure, Teratology studies, Office of Pesticide Programs, Washington DC
- USEPA (Environmental Protection Agency) (1991) Guidelines for developmental toxicity risk assessment. December 5, 1991. *Fed. Reg.* 56:63798-63826.
- USEPA (Environmental Protection Agency) (1993a) Health effects division standard evaluation procedure, Developmental toxicity studies (Draft), Office of Pesticide Programs,
- USEPA (Environmental Protection Agency) (1993b) Health effects division standard evaluation procedure, reproductive toxicity studies (Draft), Office of Pesticide Programs, Washington DC
- USEPA (Environmental Protection Agency) (1996) Guidelines for reproductive toxicity risk assessment. October 31, 1996. *Fed. Reg.* 61: 56274-56322.
- USEPA (Environmental Protection Agency) (1998a) Health Effects Test Guidelines OPPTS 870.6300 Developmental Neurotoxicity Study. Office of Prevention, Pesticides and Toxic Substances, USEPA. EPA 712-C-98-239. 12 pp.

USEPA (Environmental Protection Agency) (1998b) Guidelines for Neurotoxicity Risk Assessment. May 14, 1998. Fed. Reg. 63:26926-26954.

USEPA (Environmental Protection Agency) (1999) Neurotoxic pesticides: availability of data call-in notice. Fed. Reg. 64: 42945-42947.